IRE Transactions

ON AUTOMATIC CONTROL

Volume AC-5 AUGUST, 1960 Number 3

Rashis

TABLE OF CONTENTS

Reprinted from Part 4 of the 1960 IRE INTERNATIONAL CONVENTION RECORD (Sessions 1 and 9)

PUBLISHED BY THE

PROFESSIONAL GROUP ON AUTOMATIC CONTROL

IRE PROFESSIONAL GROUP ON AUTOMATIC CONTROL

The Professional Group on Automatic Control is an organization, within the framework of the IRE, of members with principal professional interest in Automatic Control. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of the prescribed fee.

Annual Fee: $2.00

IRE TRANSACTIONS®
on Automatic Control

George S. Axelby, Editor, Air Arm Division, Westinghouse Electric Corp., Box 746, Baltimore, Md.

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Automatic Control, 1 East 79th Street, New York 21, New York. Respsonsibility for the contents rests upon the authors, and not upon the IRE, the Group or its members. Individual copies available for sale to IRE-PGAC members at $.55, to IRE members at $.80, and to nonmembers at $1.65.

PREFACE

Since Professional Group members no longer receive free copies of the CONVENTION RECORDS for the March International Convention and WESCON, the Administrative Committee of the Professional Group on Automatic Control has established a policy of reprinting the PGAC portions of these CONVENTION RECORDS as special issues of the PGAC TRANSACTIONS insofar as budget limitations will permit.

It should be noted, however, that the papers presented at these two conventions are not subject to the same review process which has been established for the papers presented in regular issues of the PGAC TRANSACTIONS. Regular TRANSACTIONS papers are reviewed in completed form by a review committee and returned to the author for corrections and changes, but the papers presented at the conventions are seldom available in completed form prior to the convention, and they are photo reproduced from typed copy submitted by the author for the CONVENTION RECORD. Therefore, to facilitate publication, to provide a complete record of the convention control papers, and to save costs, the completed paper is subjected to no formal review process.

—The Editor

THE INCREMENTAL PHASE PLANE FOR NONLINEAR SAMPLED DATA SYSTEMS

John A. Aseltine and Richard A. Nesbit
Space Technology Laboratories, Inc.
University of California
Los Angeles, California

Abstract

An analysis technique for nonlinear sampled data systems is developed, using the incremental phase plane. This method is analogous to the phase plane method for continuous systems. A sampled data system with saturation is analyzed to demonstrate the use of the incremental phase plane in system analysis. Path tangent curves are introduced which allow the graphical solution of systems with more general types of nonlinearities. A difference equation analog of Van der Pol's equation is solved using the path tangent curves.

Introduction

Difference equations bear the same relation to sampled data systems as differential equations do to continuous ones. This suggests that there might be a counterpart to the phase plane, but based on differences, which could be applied to nonlinear sampled-data systems. This paper describes such a technique by introducing the incremental phase plane or Δ-plane. In the Δ-plane we plot the first difference (as opposed to the first derivative) against the system variable. It can be applied to sampled systems in a way similar in some respects to the application of phase plane to continuous systems.

Before beginning the development of the method it is well to call attention to other methods proposed for nonlinear sampled systems. The conventional phase plane was first used by Alter and Helstrom[1] to analyze oscillatory systems. Kalman[2] uses a "normal mode" phase plane in which system variables are plotted after a coordinate transformation. More recently Mullin[3] has combined conventional phase plane with switching lines indicating sampling times. Schmitt[4] treats the nonlinearity as an equivalent gain in his studies. The first three of these methods are based on the conventional phase plane. That is, they make use of derivatives, rather than differences, in their plots.

In this paper we shall begin by showing the genesis of the difference equation describing the sampled system. Next, the Δ-plane will be introduced and Δ-plane trajectory equations defined and obtained. As an example, a system with saturation will be analyzed--an example of interest since the other graphical methods[2,3] have had difficulty treating it. Finally path tangent curves will be introduced and used to study a Van der Pol type sampled system.

The Δ-plane method is applicable to second-order systems containing a variety of nonlinearities. In addition to saturation discussed in this paper, the method has been applied to systems with bang-bang and other types of control for a number of plant transfer functions.[5] We now proceed to develop the steps leading to the Δ-plane.

Genesis of the Difference Equation

Although the difference equation is the natural expression of behavior of a sampled system, we arrive at it in a round about way. This is due in part to the way in which sampled-data analysis evolved. We will here summarize the steps briefly.

Figure 1 shows a linear system described by transfer function G(s). At its input and output are samplers which produce impulse functions periodically, each having an area equal to the value of the sampled variable at that instant.

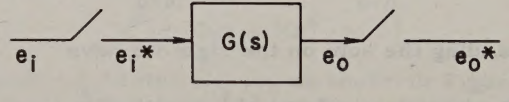


FIGURE 1. SAMPLED LINEAR SYSTEM

The samplers are shown by the switches in the figure. For convenience we shall normalize the sampling period to unity. An input is applied at time zero. The sampled input, then, is given by

$$e_i^*(t) = \sum_{n=o}^{\infty} e_i(t)\ \delta(t-n) \qquad (1)$$

The $\mathcal{L}$ transform of this is

$$e_i^*(s) = \sum_{n=o}^{\infty} e_i(n)\ e^{-ns} \qquad (2)$$

If we now replace e^s by z we have the $\mathcal{Z}$-transformed input

$$E_i(z) = \sum_{n=o}^{\infty} e_i(n)\ z^{-n} \triangleq \mathcal{Z}\left[e_i(n)\right] \qquad (3)$$

Note that in (3) we have also defined the $\mathcal{Z}$ transform.*

*See reference 6 for details of sampled systems; and reference 7, Chap. 16 for $\mathcal{Z}$ transforms.

Next, the $\mathcal{L}$-transformed output is given by

$$E_o(s) = G(s)\, E_i^*(s) \tag{4}$$

Without going into the details of the derivation which can be found elsewhere,[6] we have for the $\mathcal{Z}$-transform of the output

$$E_o(z) = G(z)\, E_i(z) \tag{5}$$

where

$$G(z) = \left\{ \mathcal{L}\,[h^*(t)] \right\}_{z=e^s} \tag{6}$$

in which h^* is the sampled impulse response. We now have means of expressing the transforms of sampled input and output of a linear system.

Let us turn now to a question concerning the $\mathcal{Z}$ transform defined by (3). If we were to begin a table of operation-pairs, we might start with

$e(n)$	$E(z)$
$e(n+1)$	?

where the question arises regarding the operation in the z-domain corresponding to shifting in n. From (3) we write

$$\mathcal{Z}[e(n+1)] = \sum_{n=0}^{\infty} e(n+1)\, z^{-n} = z \sum_{n=0}^{\infty} e(n+1)\, z^{-(n+1)} \tag{7}$$

Expanding the sum on the right we have

$$\begin{aligned} \mathcal{Z}\,[e(n+1)] &= z\,[e(1)\, z^{-1} + e(2)\, z^{-2} + \ldots] \\ &= z\,[E(z) - e(0)] \end{aligned} \tag{8}$$

and we can fill in the table

$e(n)$	$E(z)$
$e(n+1)$	$z\,E(z) - z\,e(0)$

We see that multiplication by z corresponds to shifting in n in the same way that multiplication of a Laplace transform by s corresponds to differentiation.

Returning now to (5) we write

$$\frac{E_o(z)}{E_i(z)} = G(z) = \frac{A_q z^q + A_{q-1} z^{q-1} + \ldots + A_o}{b_r z^r + b_{r-1} z^{r-1} + \ldots + b_o} \tag{9}$$

where we have assumed that $G(z)$ is a rational fraction. Multiplying (9) out, we have

$$\begin{aligned} &(b_r z^r + b_{r-1} z^{r-1} + \ldots + b_o)\, E_o(z) \\ &= (A_q z^q + A_{q-1} z^{q-1} + \ldots A_o)\, E_i(z) \end{aligned} \tag{10}$$

and, finally, using the shifting property (8) and assuming initial conditions are included in the input, the difference equation results:

$$\begin{aligned} &b_r e_o(n+r) + b_{r-1} e_o(n+r-1) + \ldots + b_o e(n) \\ &= A_q e_i(n+q) + A_{q-1} e_i(n+q-1) + \ldots + A_o e_i(n) \end{aligned} \tag{11}$$

Thus, the correspondence

$$e(n+1) \;\Big|\; z\,E(z)\,,\quad e(0) = 0 \tag{12}$$

and the one related to the first difference

$$\Delta e(n) \;\Big|\; (z-1)\,E(z)\,,\quad e(0) = 0 \tag{13}$$

where

$$\Delta e(n) \triangleq e(n+1) - e(n) \triangleq (\mathbf{E}-1)\, e \tag{14}$$

become the keys to the difference-equation description of a sampled system.

As an example, consider the double integrator shown in Figure 2.

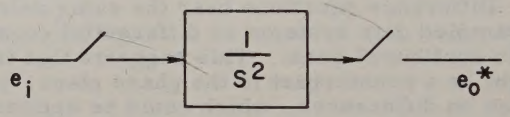


FIGURE 2. DOUBLE INTEGRATOR

The z block diagram is shown in Figure 3 with the corresponding z transfer function.[6]

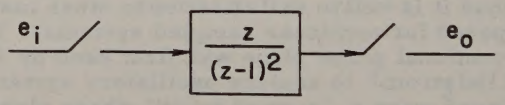


FIGURE 3. z BLOCK DIAGRAM OF DOUBLE INTEGRATOR

Now we write

$$\frac{E_o(z)}{E_i(z)} = G(z) = \frac{z}{(z-1)^2} \tag{15}$$

or

$$(z-1)^2\, E_o(z) = z\, E_i(z) \tag{16}$$

and, using (12) and (13), the system difference equation results:

$$\Delta^2 e_o(n) = e(n+1) \tag{17}$$

The Δ-Plane

The incremental phase plane, or Δ-plane, is shown in Figure 4. Because of the definition of the Δ operator (14) the ordinate at any point is equal to the difference in abscissa to the next sample point.

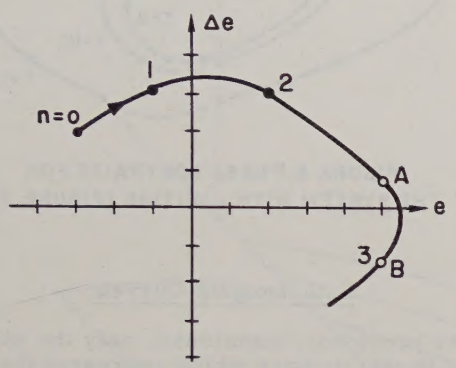


FIGURE 4. Δ—PLANE

Each point of a solution specifies $\underline{x}_n$ and $\Delta\underline{x}_n$. The value of $\underline{x}_{n+1}$ is obtained from $\underline{x}_{n+1} = \underline{x}_n + \Delta\underline{x}_n$. The value of $\Delta\underline{x}_{n+1}$ can be determined by either of two methods. One method uses trajectories, the other uses path tangent curves.

A trajectory is defined to be a curve which passes through all points of a solution that has any point on the curve.

It should be pointed out that the values of sampled system variables between sampling instants are not defined along the trajectory, which only connects values at the sampling instants. When an ambiguity arises, as between points A and B in the figure, the furthest point reached along the trajectory is the correct one.[5]

Trajectories

Let us write the second-order difference equation

$$\Delta^2 x = K \tag{18}$$

as two first-order equations

$$\begin{aligned} \Delta y &= K \\ \Delta x &= y \end{aligned} \tag{19}$$

The equation of the trajectory for the system described by these equations is

$$y^2 - K(2x+y) = C \tag{20}$$

as can be verified by differencing (20). [A method of direct solution of (19) in the form of (20)--that is, without first knowing the solution to the paired equation[2]--is not known to the authors.] The trajectory (20) is given by a family of parabolas offset from the x axis, one of which is shown in Figure 5.

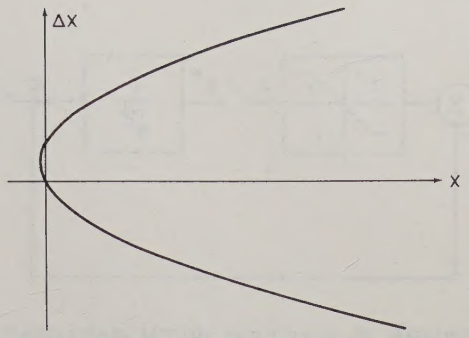


FIGURE 5. TRAJECTORY FOR $\Delta^2 X = 1$

Another case of interest is the equation for a purely oscillatory system with frequency ω_o:

$$\Delta^2 x + 2(1-\cos\omega_o)\Delta x + 2(1-\cos\omega_o)x = 0 \tag{21}$$

which we write

$$\Delta y + K(y + x) = 0, \qquad K \triangleq 2(1-\cos\omega_o) \tag{22}$$

The trajectory equation is

$$y^2 + Kxy + Kx^2 = C^2 \tag{23}$$

a family of rotated elipses as shown in Figure 6 for K = 1. Equation (23) can be verified as a solution of (22) by differencing.

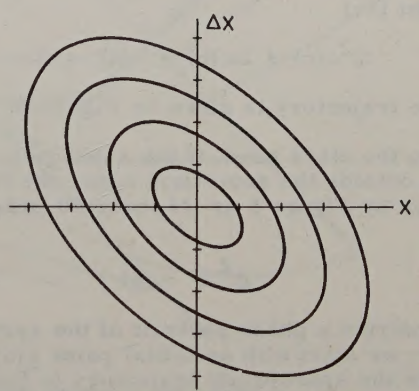


FIGURE 6. TRAJECTORIES FOR $\Delta^2+\Delta X+X=0$

These trajectories will be used in the example treated in the next section.

A System with Saturation

We shall illustrate the application of the Δ-plane method by considering the system with saturation shown in Figure 7.

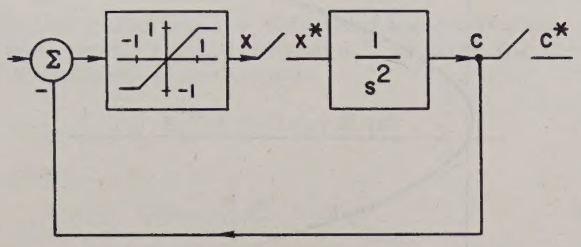


FIGURE 7. SYSTEM WITH SATURATION

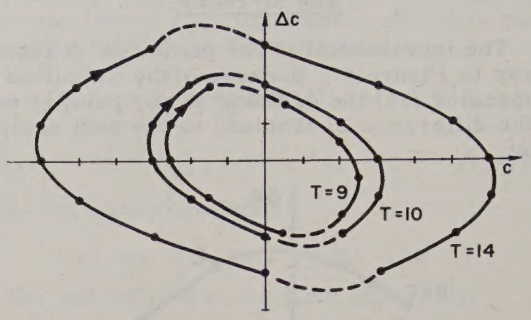


FIGURE 8. PHASE PORTRAITS FOR THE SYSTEM WITH LIMITING (FIGURE 7)

From (17) we write (assuming the input zero)

$$\Delta^2 c(n) = x(n+1) \tag{24}$$

But x, the output of the limiter has a value depending on c, so

$$\Delta^2 c(n) = \begin{cases} 1 & c(n+1) < -1 \\ -c(n+1) & |c(n+1)| < 1 \\ -1 & c(n+1) > 1 \end{cases} \tag{25}$$

In other words, if the subsequent value of c lies between ± 1, the equation is

$$\Delta^2 c(n) + c(n+1) = 0 \tag{26}$$

or from (14)

$$\Delta^2 c(n) + \Delta c(n) + c(n) = 0 \tag{27}$$

and the trajectory is given by Figure 6.

On the other hand, if the subsequent value of c lies outside the saturation zone, the trajectory is given by Figure 5 or its inverted image, since then

$$\Delta^2 c = \pm 1 \tag{28}$$

To construct a phase portrait of the system behavior, we start with an initial point $c(o)$, $\Delta c(o)$, and use the appropriate trajectory to find subsequent points. At each point we determine whether the following one lies within a different region as defined by (25). If a new region is entered, the trajectory is changed accordingly.

Several phase portraits for this system are shown in Figure 8. For the initial conditions chosen, limit cycles of periods 9, 10, and 14 result. The dased parts of the portraits correspond to trajectories chosen from Figure 6. The other parts come from Figure 5 or its inverted image.

Path Tangent Curves

As previously mentioned, only the discrete points in the Δ-plane which represent the sampled quantities are significant, and not the points in between. If the consecutive sample points are connected with line segments, the resulting polygonal line will be referred to as a solution path. The slope of the path segment is $\Delta y/\Delta x$ where $y = \Delta x$.

A method for tracing solution paths for equations which may have continuous nonlinearities is obtained by considering the similar differential equation. The ordinary phase plane trajectories of this similar equation are tangent to the segments of the solution path. The similar equation is obtained by replacing Δx by $\dot{x}$ and $\Delta^2 x$ by $\ddot{x}$. The slope of the ordinary phase plane trajectory is dy/dx and the slope of the path segment is $\Delta y/\Delta x$. The usual methods of obtaining the phase plane trajectories may thus be used to study the sampled system. For the previous example, the similar equations are:

$$\begin{aligned} \ddot{c} &= \pm 1 && |\dot{c} + c| > 1 \\ \ddot{c} + \dot{c} + c &= 0 && |\dot{c} + c| < 1 \end{aligned} \tag{29}$$

The path tangent curves are sketched in Figure 9 with a typical path indicated.

A path is obtained by continuing along the tangent of the path tangent curve through the starting point until the x-coordinate has changed by the Δx indicated at the starting point. This procedure becomes indeterminate for starting points where $x = 0$. Along the x-axis, the original equation is used to determine Δy.

For systems which switch on some boundary in the Δ-plane, the path tangent curves switch on the same boundary and do not overlap as the trajectories may do. The singular points of the similar differential equation are also singularities in the Δ-plane. Perturbation about these singularities gives information about their stability.

The system shown in Figure 10 is described by the difference equation

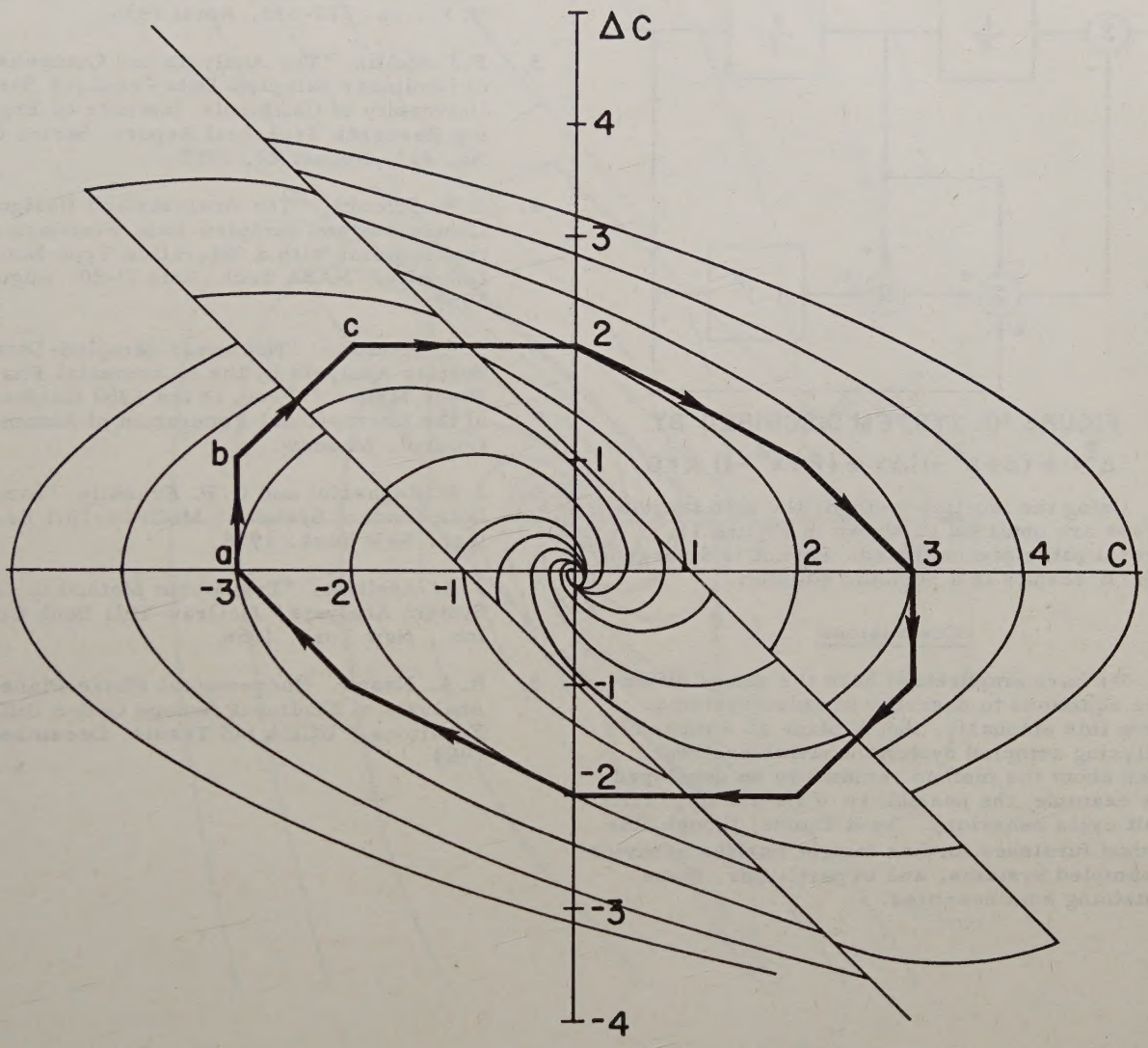


FIGURE 9. PATH TANGENT SOLUTION FOR SYSTEM IN (FIGURE 8)

$$\mathbf{E}^2 x + (x^2 - 1)\,\mathbf{E}x + x = 0 \qquad (30)$$

or

$$\Delta^2 x + (2+x^2-1)\Delta x + (2+x^2-1)x = 0 \qquad (31)$$

The similar differential equation is

$$\ddot{x} + (2+x^2-1)\dot{x} + (2+x^2-1)\,x = 0 \qquad (32)$$

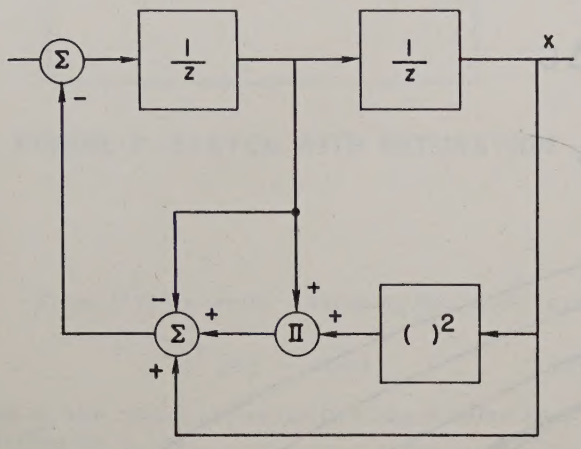


FIGURE 10. SYSTEM DISCRIBED BY $\Delta^2 x + (2+x^2-1)\Delta x + (2+x^2-1)\,x = 0$

Using the isocline method, the path tangent curves are obtained as shown in Figure 11. Several paths are indicated. Path A is divergent; path B results in a periodic solution.

Conclusions

We have emphasized here the use of difference equations to describe sampled systems. From this emphasis, the Δ-plane as a means of analyzing sampled system behavior evolved. Much about the method remains to be developed (for example, the possibility of its use to predict limit cycle behavior). As it stands, though, the method furnishes further insight into the behavior of sampled systems, and in particular, those containing nonlinearities.

References

1. Alter and Helstrom, "Phase-Plane Representation of Sampling Servomechanisms," Research Report R-94410-14-B, Westinghouse Research Labs., E. Pittsburgh, Pa., Sept. 21, 1953.

2. R. E. Kalman, "Nonlinear Aspects of Sampled-Data Control Systems," Proc. Symp. on Nonlinear Circuit Analysis, Poly. Inst. Brooklyn, N. Y., pp. 273-313, April 1956.

3. F.J. Mullin, "The Analysis and Compensation of Nonlinear Sampled-Data Feedback Systems," University of California, Institute of Engineering Research Technical Report, Series 60, No. 211, August 22, 1958.

4. S. F. Schmitt, "The Analysis and Design of Continuous and Sampled-Data Feedback Control Systems with a Saturation Type Nonlinearity," NASA Tech. Note D-20, August 1959.

5. J. A. Aseltine, "Nonlinear Sampled-Data System Analysis by the Incremental Phase Plane Method," Proc. of the 1960 Conference of the International Federation of Automatic Control, Moscow.

6. J. R. Ragazzini and G. F. Franklin, "Sampled-Data Control Systems," McGraw-Hill Book Co., Inc., New York, 1958.

7. J. A. Aseltine, "Transform Method in Linear System Analysis," McGraw-Hill Book Co., Inc., New York, 1958.

8. R. A. Nesbit, "Incremental Phase Plane Analysis of Nonlinear Second Order Difference Equations," UCLA MS Thesis, December 1959.

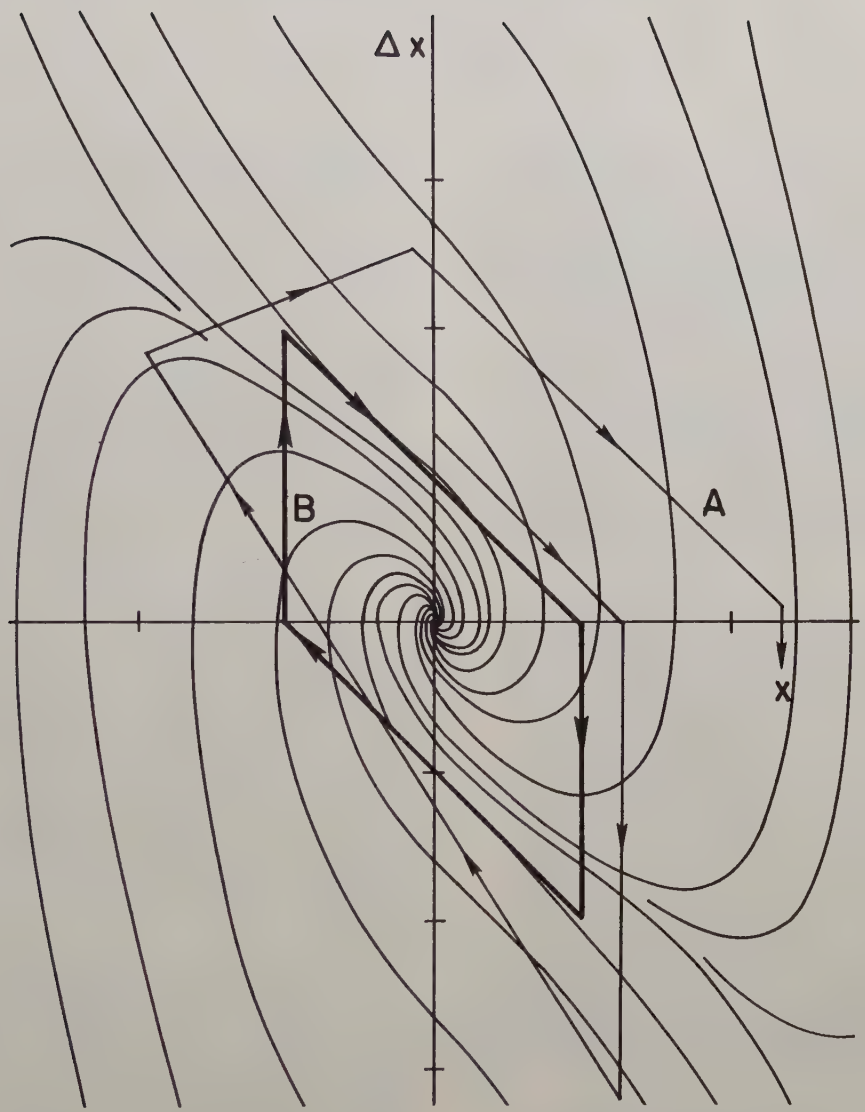


FIGURE II. PORTRAIT OF $\Delta^2 x + [2 + (x^2 - 1)]\,\Delta x + [2 + (x^2 - 1)]x = 0$

ON THE EXISTENCE AND UNIQUENESS OF THE OPTIMAL MULTIVARIABLE SYSTEM SYNTHESIS

Mihajlo Mesarović
Case Institute of Technology
Cleveland, Ohio

Many problems in the study of multivariable systems are not of the same nature as the problems normally associated with single variable systems. This has interesting consequences when solving practical problems of the synthesis. For instance, when in the synthesis of single variable systems the available analytical procedures cannot provide the solution one can simulate the problem on the computer (analog or digital) and obtain numerical solution for the particular situation at hand. However, in many problems in synthesis of the multivariable systems the solution itself depends upon the way of representing the system. Such a situation appears in synthesis of the cross controllers when the resulting changes in the interrelations of the closed loop system should be taken into account.

We shall consider some specific properties of the multivariable system synthesis which are valid for a broad class of problems rather than look for the closed-form solution of some special problems. Actually in the absence of an adequate closed form solution one can look for a numerical solution but in doing so one should be aware of the general properties of the problem at hand. In a monograph to appear soon one can find more detailed discussion on the specific properties of the multivariable systems.[1] Here we shall be concerned with the problem of the feasibility of synthesis. More accurately we shall consider the following problem: Given a set of input variables and a set of the corresponding desired output variables which kind of systems might offer a solution and whether the resulting system is the only one to satisfy the requirement of the synthesis.

In the conventional formulation of the synthesis problems in single variable systems, the existence and uniqueness of the solution ultimately depends upon the possibility of finding a unique solution of the equations defining the optimal conditions. However, if the synthesizing procedure does not offer a unique solution it is important to find the reason for that. In general, there are three steps in which the optimizing procedure can fail:

1) The equation for optimal behavior may have no solution, i.e. even the ideal optimal system cannot be found.

2) The ideal optimum can be found but not a satisfactory approximation of the optimum in a restrictive class of functions such as a class of physically realizable functions technically realizable functions (in a given sense), etc.

3) One cannot synthesize the system with the hardware at hand, even if its behavior belongs to a restrictive class of functions.

The problems in each of the above three steps become even more complex in the case of a multivariable system. In addition there is one step more in which the unique synthesis of the multivariable systems can fail: namely, the optimal conditions can fail to offer a large enough set of equations, the solution of which should specify the optimal system. This step precedes the above mentioned, and in what follows we shall be discussing relating problems.

Binary Representation of Systems

Before further discussion let us consider the possibilities for a more precise definition of the synthesis task. In the synthesis of the system mainly two kinds of problem may appear: (1) The inputs and desired outputs are given and one should synthesize the system. This class of problem could be termed filter-problem. (2) In addition to the desired outputs and inputs the system (or process) is given and one should synthesize another set of inputs, often termed manipulated variables in the control type of problems.

We shall show in this section in what sense these two kinds of problem are analogous. To this end we shall use binary representation of the system. In this kind of representation every function among the variables is represented as a binary operation, i.e. it depends upon two variables (two inputs) and produces one variable (one output). The linear function can now be represented by convolution as binary operation which, as two inputs, has the actual input and weighting function of the element. A static nonlinear element can be represented in many different ways. It can be represented by multiplication as binary operation with one of the multiplicants (which is now time variant) determining the kind of nonlinearity. Multiplication is not, in fact, always convenient because it occasionally requires infinite values for one of the multiplicants. As a better way of representing the nonlinearity the potential with time variant exponent can be used. The memory-type nonlinearities might require many cascading binary operations for an adequate representation.

The choice of the kind of binary operation is to some extent arbitrary. Preferably the operation itself should not include quantitative information, or at least no information subjected to synthesis. In other words the binary operation should specify only the kind of operation to be performed while all associated data and variables should be represented as inputs. A linear system e.g. can be represented by the binary structure of the system which includes only convolution and addition as binary operations and two sets of inputs. One set of the inputs consists of weighting functions of all elements. They could be regarded as generalized parameters and will be referred to as behavioral inputs. The other set consists of the input variables of the system and for the sake of distinction they will be termed environmental inputs. A general representation of the system is as in Fig. (1). The binary structure includes many different binary operations.

The problem in the synthesis that we want to consider can now be formulated in the following way: Given the binary structure of the system and one set of inputs (environmental as well as behavioral inputs can be included in this set) the remaining set of inputs should be synthesized.

A General Form of Necessary Conditions For Optimal Multivariable Synthesis

To discuss problems in the first step of synthesis we are in need for necessary conditions under rather general circumstances. To this end we shall take as the performance criterion, the functional of the form

$$J = \int_{t_1}^{t_2} F\left[y_1(t), \ldots\ldots y_n(t)\right] dt \qquad (1)$$

where: F is a nonlinear function of all the outputs $y_1, \ldots\ldots y_n$.

The system is assumed to be described by the equations

$$G_j\left[z_1, \dot{z}_1, \phi(z_1), \ldots z_k, \dot{z}_k \phi(z_k), \ldots \ldots y_1, \dot{y}_1, \phi(y_1) \ldots\ldots y_n, \dot{y}_n, \phi(y_n)\right] = 0 \qquad (2)$$

where $z_1 \ldots z_k$ are inputs to the binary structure which should be synthesized $\phi(y_1) \ldots \ldots \phi(y_n)$ are functionals which describe the linear parts of the system. The inputs to the system $x_1, \ldots x_n$ are supposed to be known deterministic time functions and therefore are not explicitly shown in Eq. (2).

Consider the case where the optimization should be performed only with respect to one of the inputs, say, z_j. The calculus of variation technique gives for the optimizing condition

$$\sum_k \lambda_k T_{G_k} * G_k(z_j) = 0 \qquad (3)$$

where the operator T defined by

$$\frac{\partial K(x)}{\partial x} - \frac{d}{dt}\left(\frac{\partial K(x)}{\dot{x}}\right) + \frac{\partial K(x)}{\partial \phi}\left(\frac{\partial \phi}{\partial x}\right) = T * K \qquad (4)$$

was introduced. Function K in Eq. (4) can be either G or F.

where: $\lambda_k = \frac{\Delta k}{\Delta}$

Determinants Δ_k and Δ do not depend upon the input z_j which is to be synthesized.

$$\Delta = \left\{T_{G_j} * G_j(y_k)\right\} \qquad (5)$$

Another interesting form for the necessary condition of the optimum can be obtained if the behavior of the system is given by the equations.

$$y_j = G_j\left[z_{j_1}, \dot{z}_{j_1}, \phi(z_{j_1}), \ldots z_{j_k}, \dot{z}_{j_k}, \phi(z_{j_k}), y_1, \dot{y}_1, \phi(y_1), \ldots y_{j-1}, \dot{y}_{j-1}, \phi(y_{j-1}), y_{j+1}, \dot{y}_{j+1}, \phi(y_{j+1}) \ldots y_n, \dot{y}_n, \phi(y_n)\right] \qquad (6)$$

The optimal condition then becomes

$$\left[\sum_1 \frac{\partial F}{\partial y_1} \frac{\partial y_1}{\partial y_j}\right] \frac{\partial y_j}{\partial z_{j_p}} = 0 \qquad (7)$$

Proceed now to consider the case with the random inputs. Assume that the system is linear and adopt the cononical structure for it. The equations describing the behavior of the process are

$$Y_j(t) = \sum_k \int_0^\infty P_{jk}(\tau)\, X_k(t - \tau)\, d\tau \qquad (8)$$

Take as a performance measure a function of all the outputs $Y_1, \ldots Y_n$, and reference variables $Y_{1r}, \ldots Y_{nr}$

$$e = e\left[Y_1, \ldots Y_n, Y_{1r}, \ldots Y_{nr}\right] \qquad (9)$$

Expanding the performance criterion in Taylor series and taking only the first members one obtains

$$e = e(\bar{y}_1, \ldots \bar{y}_n, \bar{y}_{1r}, \ldots \bar{y}_{nr}) + \sum_{1}^{n+nr} \frac{\partial e(\bar{y}_1, \ldots \bar{y}_n, \bar{y}_{1r}, \ldots \bar{y}_{nr})}{\partial \bar{y}_1} \left[Y_1 - \bar{y}_1 \right] \qquad (10)$$

where bar denotes mathematical expectation of the respective variable.

The mean square of the performance variable is now

$$\bar{e}^2 = \sum_{j}^{n+nr} \sum_{k}^{n+nr} \frac{\partial e}{\partial \bar{y}_j} \frac{\partial e}{\partial \bar{y}_k} \phi_{jk}(0) \qquad (11)$$

where ϕ_{jk} is the cross-correlation function.

As given by Eq. (11) the final performance measure is a linear function of various correlation function of the output variables and therefore has no extremum in respect to them. For greater generality assume that the performance measure $\bar{e}^2$ is a nonlinear function of correlation functions

$$\bar{e}^2 = F\left[\ldots \phi_{jk}(0) \ldots \right] \qquad (12)$$

The system can now be represented by a deterministic model which has equivalent behavior when measured by the performance criterion (12). The equivalent deterministic model has $4n^2$ outputs and $4n^2m^2$ inputs. The problem of synthesis is now in determining the behavioral inputs, i.e. the transfer functions of the elements.

Existence and Uniqueness Numbers of Multivariable Synthesis

We shall now discuss the feasibility in formulation of optimal conditions. As was mentioned in section 1 this is the first step in the multivariable synthesis in which either existence or uniqueness of the synthesis can fail to be obtained.

The ultimate goal of the synthesizing problem, as defined here, is to determine a set of inputs to the system. However to achieve a total optimum according to Eq. (1) only one subset of the inputs could be sufficient. The inputs belonging to this subset will be called the bounded inputs because they are completely specified with the synthesizing procedure. The remaining inputs of the set will be termed free because no matter how one chooses them the bounded inputs might assure a totally optimal system.

The existence and uniqueness in formulation of the optimal conditions can now be determined in an extremely simple way by considering only two numbers. The first one, termed existence number, represents the difference between the number of the outputs and the number of the bounded inputs.

$$D = r_d - n \qquad (13)$$

Its sign indicates the existence of the set of the conditions which specify the optimal system. The second number, termed the uniqueness number, represents the difference between the total number of the elements in the input set and the number of the bounded inputs or, in other words, it represents the number of elements in the set of the free inputs.

$$R = r - r_d \qquad (14)$$

The usefulness of the concepts of existence and of uniqueness number comes from the fact that in a given situation they reflect the kind of the systems structure and at the same time the necessary information one should have for the optimal synthesis.

For the linear systems one has the conditions (7). Taking into account that for linear systems one has

$$\frac{\partial y_j}{\partial z_{j_p}} \neq 0 \qquad (15)$$

the existence and uniqueness numbers become

$$D = 0;\ R = n(m - 1) > 0 \qquad (16)$$

Therefore the conditions for the total optimum of the linear system exist but are not unique, i.e. many different conditions may be formulated resulting in many different systems all having some behavior (as measured by the given performance criterion).

For nonlinear systems, one obtains from Eq. (3).

$$D = 0 \quad R > 0 \qquad (17)$$

The optimal conditions exist and might also be unique. The nonlinearity in the structure of the system can be measured just by R. Whenever one has $R > 0$, i.e. there is at least one of the inputs free, the system is not completely nonlinear in structure. The degree of the nonlinearity in structure could be measured by the value of R.

When the given inputs to the system are of random nature, one obtains according to Fig. (2)

$$D = n(m - 3n) \qquad (18)$$

D is now negative and the set of conditions for a totally optimal system does not exist.

This comes from the fact that a deterministic equivalent model has more outputs than inputs available for the synthesis. The existence of the optimal synthesis depends now upon the correlation of various outputs and reference variables, e.g. if every output is correlated only with the corresponding reference variable one has

$$D = n(m - 2) \qquad (19)$$

The existence of the optimal synthesis depends now upon the number of the variables associated with the system. One can obtain the optimal system only if $n = 2$. The increase in the number of variables of the system enables an easier synthesis.

R and D reflect the basic properties of the system in an utmost simple way and therefore they should be considered as characteristics of the system as e.g. the numbers of inputs and outputs. Therefore for the synthesis of a multivariable system one should have in advance:

1) Four characteristic numbers of the system, n, m, R and D.

2) A set of inputs and a set of the respective desired outputs.

3) The specification of the binary structure is perferable and in many practical cases presents no limitations whatsoever.

Additional Specifications for Synthesis of Totally Optimal Systems

For systems, $R > 0$, additional specifications are necessary if the synthesis is to be unique. This is in sharp contrast with the properties of the single variable problems.

Two ways might be offered for specification of the additional conditions.

1) One could specify in more detail the structure of the system as e.g. by requiring a given kind of interactions inside the system.

2) One can specify a large enough number of sets of inputs and desired outputs. One has the possibility then to cover an entire class of input functions offering a more realistic synthesis of the system.

The synthesis with the additional specifications presents a basic and very interesting problem in the multivariable study but is out of the scope of this presentation.

[1] M. Mesarović, "Foundations for a Control Theory of Multivariable Systems", Technology Press, John Wiley, 1960.

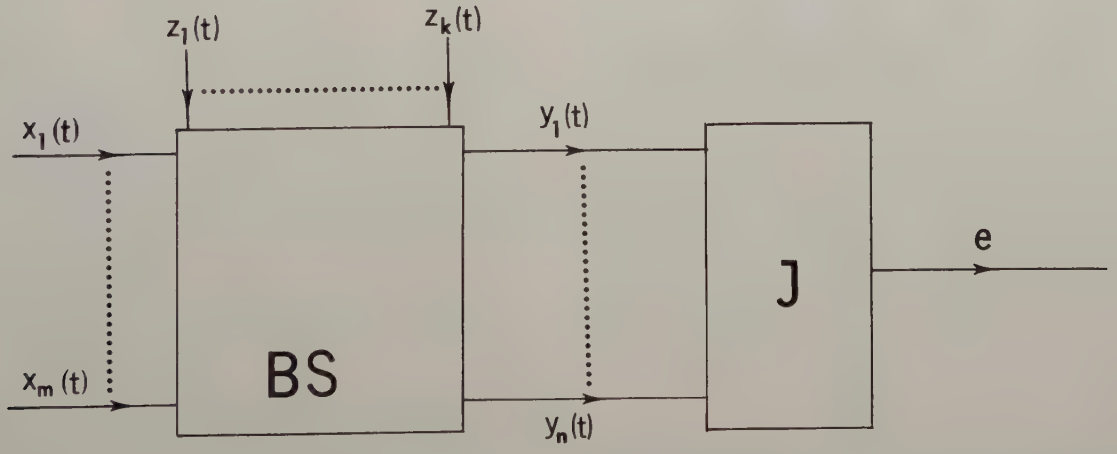


Fig. 1. Binary representation of a multivariable system. BS = binary structure, Z_j = environmental inputs, X_j = behavioral inputs, e = performance measure.

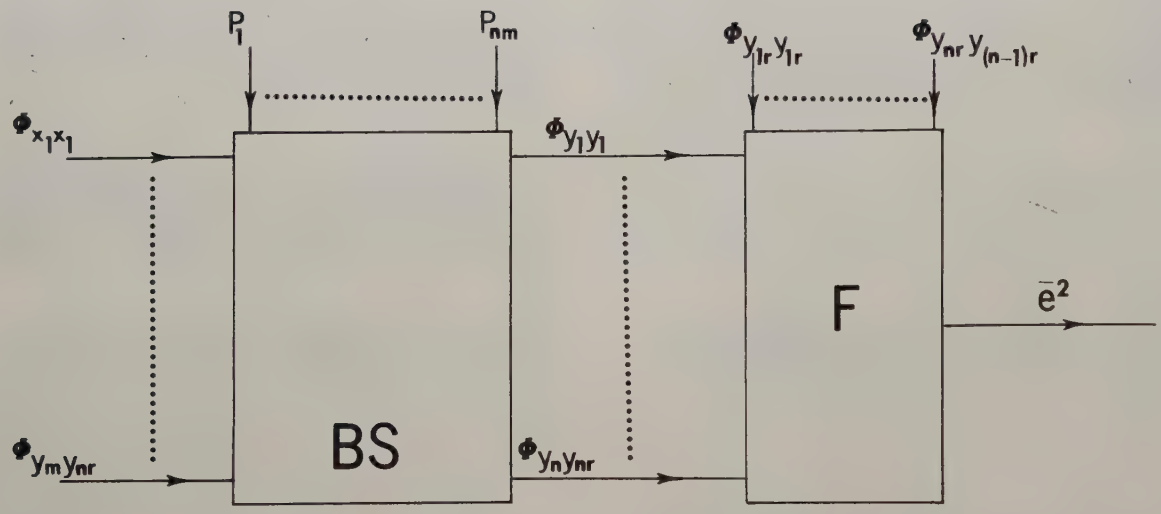


Fig. 2. Deterministic analogue of a multivariable system with random inputs.

ON OPTIMAL AND SUBOPTIMAL POLICIES IN THE CHOICE OF CONTROL FORCES FOR FINAL-VALUE SYSTEMS

by

Masanao Aoki
Department of Engineering
University of California, Los Angeles

Summary

A simple control system whose state variable x_n is described by the difference equation (1) is considered.

$$x_{n+1} = \alpha x_n + r_n + v_n, \quad n = 1, 2 \ldots N, \quad 0 < \alpha < 1 \qquad (1)$$

where v_n is the control force of the system and r_n is the Bernoulli random noise with the probability parameter p.

If $\phi(x_N)$ is the performance index of the final-value system, then the problem is to find a sequence of control forces $\{v_n\}$, $n = 1, 2 \ldots N$, which minimizes the expected value of $\phi(x_N)$. An optimal sequence of v_n is determined by solving a recurrence relation on n of the criterion function $h_n(x;p)$ defined as follows:

$$h_n(x;p) = \underset{v_1}{\text{Min}} \; \underset{v_2}{\text{Min}} \ldots \underset{v_n}{\text{Min}} \; E\left[\phi(x_n)\right]$$

= the minimum of the expected value of $\phi(x_n)$ obtainable in the n-stage control process by employing an optimal policy, starting from the initial state variable x when the probability parameter of the random disturbance is p.

The recurrence relation is obtained by the usual application of the principle of optimality of dynamic programming technique.

It is proved that $h_n(x;p) = h_n(-x; 1-p)$ holds and this fact is used in reducing by half the amount of computation when it is necessary to solve the recurrence relation numerically.

If the values of control variable $v_n(x;p)$ are restricted to 1 and -1 as in contactor servo systems, the boundaries between 1 and -1 control forces become too complicated to be determined analytically except for few special cases.

By solving $v_n(x;p)$ computationally for the $\phi(x_N) = x_N^2$ case, it is seen that $v_n(x;p)$ agrees with $v_1(x;p)$ fairly well for all $n > 1$. $v_1(x;p)$ is the optimal control force when there remains only one chance of exerting control forces. Hence, the suboptimal policy of always applying control forces as if only one more error correction is possible may be expected to be close to the optimal policy in minimizing $E(x_N^2)$. The control force $v_1(x;p)$ is linear in x and p and is a much more simple function to mechanize. This suboptimal policy is tried by the Monte Carlo method and found to be only slightly inferior to the optimal policy.

The behavior of x_n is investigated by assuming the control forces are given by the suboptimal policy. This approximate analysis should be good in view of the agreements between the optimal and suboptimal policies.

It is important to realize that if the adoption of a suboptimal policy results in a simplified mechanization of the optimal control forces, with only a slight reduction in the system performance, then the suboptimal policy might be optimal in a certain enlarged performance index.

What seems to be the most desirable approach to engineering problems is a unified or well-integrated one where the analytical and computational algorithms are used to supplement each other. In this paper, an attempt is made to illustrate the point, presenting at the same time a new approach to analysis and synthesis of a certain class of control systems.

Introduction

Time behavior of a control system can be regarded as successive transformations on a vector, called the state vector, which represents the state of the control system as a vector in phase space.[1,2,3]

Types of transformations to be applied are determined by specifying the control vectors as functions of the present state vector and time. Thus, if the time is assumed to advance in discrete steps, a problem of determining proper sequences of control vectors, given the criterion

of performance, can be formulated as a problem of multi-stage decision processes. The decision at each stage is a proper choice of the control vector from the domain of control vectors which may or may not be known in advance. That domain is, however, assumed not to change with time.

Functional equation techniques of dynamic programming are well suited to treatment of control processes as multi-stage decision processes.

By solving a functional equation obtained from the application of the principle of optimality to the control process in question, a sequence of optimal control vectors is obtained. This sequence is called an optimal policy. Except in a few isolated cases, analytical solutions of functional equations are generally not available. Thus, in order to obtain an optimal policy, it is often quite necessary to resort to some computational techniques to solve the functional equation numerically or to solve the functional equation in some approximate iterative fashion and make the sequence of approximate solutions converge to the true solution.

Even in numerical solutions of a functional equation, because of the limited capabilities of existing digital computers, it may be feasible to solve the equation only approximately. Thus depending on the complexities of functional equations, it often becomes necessary to investigate a sequence of non-optimal decisions. It is called a suboptimal policy.

Even when an optimal policy is available, we may be interested in some policy slightly inferior to the optimal policy, if the adoption of this suboptimal policy results in vastly simpler mechanization of control devices in a control system.

In the following, this concept of optimal policy versus suboptimal policy is pursued by describing the situtation where a stochastic control process serves as an approximate solution to a certain adaptive process and further by investigating a certain simpler control policy in the stochastic process which serves as a suboptimal policy. Concepts of stochastic and adaptive processes will be made precise later. Roughly, in stochastic processes, we are given distribution functions of random variables, whereas in adaptive processes, relevant distribution functions are imperfectly known, making it necessary to gather additional information on distribution functions from the history of the observations on random variables.

Adaptive Control Processes

Distribution functions for random variables affecting the performance of time-discrete control processes are assumed to be members of a certain parameter space, which is assumed to be discrete and finite. A random variable R, therefore, has the distribution function of the form $F(r, \alpha)$ where

$$p[\alpha = \alpha_i] = z_i, \quad z_i \geq 0, \qquad i = 1, 2 \ldots K;$$

$$\sum_{i=1}^{k} z_i = 1$$

Let us associate the states of nature with each α_i, $i = 1, 2 \ldots K$. They are assumed not to change with time.

If $z_i = 1$ for some i, then the control system is known to be in the i^{th} state of nature. This is a stochastic situation since the distribution function of the random variable R is given by $F(r, \alpha_i)$, a known function.

If $0 < z_i < 1$ for all $i = 1, 2 \ldots K$, then the control system is adaptive. Depending on the history of observations on R, the estimates z_i are assumed to be modified by the Bayes rule.

The adaptive model described above is a special case of the adaptive models of Bellman and Kalaba.[4]

The difference equation of a final value control system S is given by

$$x_{n+1} = T(x_n, r_n, v_n), \quad n = 0, 1 \ldots N-1 \quad (1)$$

where x_n is the state vector and v_n is the control vector of S. Let $\phi(x_N)$ be the criterion of this final value control system.

The random variable r_n is assumed to be independently and identically distributed with the distribution function $F(r, \alpha)$.

The control system S must decide to apply $\{v_n\}$ which minimizes the estimated expected value of $\phi(x_N)$, given the present state vector x, the number of the remaining decision stages n and the present estimate of the probability $P(\alpha = \alpha_i) = z_i$, $i = 1, 2 \ldots K$ of $z = (z_1, \ldots z_K)$. The estimated expected value of $\phi(x_N)$ employing the optimal sequence of v_n is expressed by $k_n(x;z)$ where x, z and n are as described above.

In the following, for the sake of brevity, we will restrict ourselves to the case of K = 2.

Let us redefine z as

$$P(\alpha = \alpha_1) = z, \quad P(\alpha = \alpha_2) = 1-z$$

Then, if we define $h_n(x;\alpha_i)$, $i = 1, 2$ correspondingly to $k_n(x;z)$, as the expected value of $\phi(x_N)$ following an optimal policy, then $k_n(x;1)$ and $k_n(x;0)$ will become equal to $h_n(x;\alpha_1)$ and and $h_n(x;\alpha_2)$, since once $z = 1$ or $z = 0$ is realized, the value of z will not change.

To obtain an optimal policy, it is necessary to solve a functional equation for $k_n(x;z)$. Under

the above conditions, however, it can be proved that $k_n(x;z)$ is concave in z,

$$k_n(x;z) \geq z\, k_n(x;1) + (1-z)\, k_n(x;0)$$

$$= z\, h_n(x;\alpha_1) + (1-z)\, h_n(x;\alpha_2) \qquad (2)$$

This equation means that knowledge of $h_n(x;\alpha_i)$, i = 1, 2 provides a lower bound on $k_n(x;z)$ without solving the functional equation for $k_n(x;z)$.

Noting the fact that α_i in $h_n(x;\alpha_i)$ is a parameter and not a part of the independent variables of the functional equation as z is in $k_n(x;z)$, it is seen that the functional equation for $h_n(x;\alpha_i)$ will in general be easier to solve. [3]

The knowledge of an optimal policy and the value of $k_n(x;z)$ is equivalent, i.e., knowing $\{k_n(x;z)\}$, all optimal policies are obtainable. On the other hand, the knowledge of an optimal policy suffices to generate $\{k_n(x;z)\}$. Therefore, the approximate knowledge on $k_n(x;z)$ as supplied by (2) may be used to provide a suboptimal policy which is constructed from our knowledge of the corresponding stochastic control process alone.

The usefulness of (2) depends entirely on how good this approximation is. We will come back to (2) later when the numerical results are discussed.

Stochastic Control Processes

In order to illustrate several points made in the previous section, let us specify the underlying difference equation (1) for S further to that given by (3).

$$x_{n+1} = \alpha x_n + r_n + v_n, \quad 0 < \alpha < 1 \qquad (3)$$

where r_n is taken to be the Bernoulli random variable

$$r_n = \begin{cases} b & \text{with probability } p \\ -b & \text{with probability } 1-p \end{cases} \qquad (4)$$

The criterion of performance is taken to be

$$E(J_N) = E\left(x_N^2 + \lambda \sum_{i=0}^{N-1} v_i^2\right) \qquad (5)$$

where the 2nd term is included to express the situation where there is a constraint on the total amount of control force in the form $\sum_{i=0}^{N-1} v_i^2 \leq V$.

Then, defining $h_n(x;p)$ to be the minimum of the expected value of J_N obtainable in the n-stage control process by employing an optimal policy, starting from the initial state variable x when the probability parameter of the random disturbance is p, the recurrence relation for $h_n(x;p)$ is given by

$$h_1(x;p) = \min_{v_o} \left[\lambda v_o^2 + p\,(x_+)^2 + (1-p)(x_-)^2\right]$$

$$h_{n+1}(x;p) = \min_{v_n} \left[\lambda v_n^2 + p\, h_n(x_+;p) + (1-p)\, h_n(x_-;p)\right] \qquad (6)$$

$$n = 1, 2 \ldots N-1$$

where $x_+ = \alpha x + b + g$, $x_- = \alpha x - b + g$

If no other constraint is imposed on v, then, as expected, $h_n(x;p)$ is quadratic in x and $v_n(x;p)$ is linear in x,

$$h_n(x;p) = (b^2 - \bar{b}^2)\left[1 + \frac{\lambda}{\lambda+1}\alpha^2 + \ldots + \frac{\lambda\,\alpha^{2(n-1)}}{\lambda + 1 + \alpha^2 + \ldots \alpha^{2(n-2)}}\right] + \frac{\lambda}{\lambda + 1 + \alpha^2 + \ldots + \alpha^{2(n-1)}} \left[\alpha^n x + \bar{b}\,(1 + \alpha + \ldots + \alpha^{n-1})\right]^2 \qquad (7)$$

$$v_n(x;p) = -\frac{\alpha^{n-1}\left[\alpha^n x + \bar{b}\,(1 + \alpha + \ldots + \alpha^{n-1})\right]}{\lambda + 1 + \alpha^2 + \ldots + \alpha^{2(n-1)}}$$

where $\bar{b} = E(b) = (2p - 1)\, b$.

However, if v_n is constrained to be

$$v_n = m \text{ or } -m, \qquad m > b \qquad (8)$$

as in a contactor servo system, then explicit expressions for $h_n(x;p)$ and $v_n(x;p)$ are no longer available. Since $\sum_{i=0}^{N-1} v_i^2 = Nm^2$, the criterion of performance can be taken simply to be

$$E(x_N^2)$$

The recurrence relation is now given by

$$h_1(x;p) = \min_{v_o = \pm m} \left[p\,(x_+)^2 + (1-p)\,(x_-)^2\right]$$

$$h_{n+1}(x;p) = \min_{v_n = \pm m} \left[p\, h_n(x_+;p) + (1-p)\, h_n(x_-;p)\right] \qquad (9)$$

$$n = 1, 2 \ldots N-1$$

Although explicit expressions for h_n and v_n are not available, it can be shown by inductive arguments that $h_n(x;p) = h_n(-x;1-p)$ holds. [3]

Due to this symmetry, the amount of computation necessary to solve $h_n(x;p)$ for a given range of x and $0 \leq p \leq 1$ is reduced by half.

The symmetry property is seen to hold even when the domain of v_n is expanded to include more than two values so long as for $v_n = a$, the corresponding value $v_n = -a$ is in the domain. ($v_n = 0$ is the special case of this).

If the value of p is regarded as the parameter α of the previous section, i.e.,

$$P\left[p=p_1\right] = z, \qquad P\left[p=p_2\right] = 1-z$$
$$1 \geq p_1 > p_2 \geq 0$$

then $k_n(x;z)$ as defined in that section will satisfy the recurrence relation

$$k_1(x,z) = \underset{v=\pm m}{\text{Min}} \left\{ \begin{array}{l} z\left[p_1\ \phi(x_+) + (1-p_1)\phi(x_-)\right] \\ + (1-z)\left[p_2\ \phi(x_+) + \right. \\ \left. (1-p_2)\ \phi(x_-)\right] \end{array} \right\}$$

$$k_{n+1}(x_1 z) = \underset{v=\pm m}{\text{Min}} \left\{ \begin{array}{l} z\left[p_1\ k_n(x_+, z') + \right. \\ \left. (1-p_1)\ k_n(x_-, z'')\right] + \\ (1-z)\left[p_2\ k_n(x_+, z') + \right. \\ \left. (1-p_2)\ k_n(x_-, z'')\right] \end{array} \right\} \tag{10}$$

where x_+ and x_- are as defined in (6), and

$$z' = \frac{p_1 z}{p_1 z + p_2(1-z)}, \quad z'' = \frac{(1-p_1) z}{(1-p_1)z + (1-p_2)(1-z)}$$

The recurrence relations (9) and (10) are solved computationally with the boundary condition that $|x_n| \leq D$ for all n for the same system parameters. This boundary condition means that there is a physical stop beyond which the system cannot go.

By comparing $k_n(x;z)$ and $z\, h_n(x;p_1) + (1-z)\, h_n(x;p_2)$ the approximation of (2) is tested. Figure (1) shows $k_n(x,z)$ as a function of z for several n and x.

Since the approximation of (2) amounts to regarding $k_n(x,z)$ to be linear in z, the degree of approximation is seen to be better for larger n.

One-Stage Suboptimal Policy

It is shown in previous sections that an optimal policy for the stochastic system (3) is obtainable by solving (9), with knowledge of $\{h_n(x;p_i)\}$ i = 1, 2 presupposed in the approximation of $k_n(x,z)$ in terms of $h_n(x;p_1)$ and $h_n(x;p_2)$.

Let us next turn to the investigation of the approximate solution of (9), i.e., the determination of a suboptimal policy of the stochastic control system (3) itself.

Figure (2) shows an optimal control variable as a function of x and n for p=0.625, $D=\frac{1}{4}$, $\alpha = 7/8$, b = 1/16, m = 9/128, $n \leq N = 12$.* Notice that the boundaries between $v_n(x;p) = +m$ and $v_n(x;p) = -m$ is not a simple straight line.

When N = 1, the optimal control vector is given by

$$v_1(x;p) = -m\,\text{sgn}\left(\alpha x + (2p-1)\, b\right) \tag{11}$$

The switching boundary, then, is a straight line given by

$$x = -(2p-1)b/\alpha \tag{12}$$

and shown in Figure (2).

If (11) is used as $v_n(x;p)$ for $n \geq 2$, then it constitutes a suboptimal policy.

From Figure (2) it is noted that for $x > 0$ the optimal and suboptimal policies agree fairly well although the agreement is not particularly good for $x < 0$.

It might be expected, therefore, that if the system starts from the initial position x = D, then the suboptimal is a fairly good approximation for the optimal policy.

To see to what extent the conjecture is correct, the Monte Carlo method was used to follow the system behavior from initial positions $x_0 = \pm D$ for 40 trials each.

With $x_0 = D$, $E(x_N^2)$ optimal = 0.00367,

$E(x_N^2)$ suboptimal = 0.00374.

With $x_0 = -D$

$E(x_N^2)$ optimal = 0.00437,

$E(x_N^2)$ suboptimal = 0.00441.

Despite the relatively similar values in $E(x_N^2)$, the suboptimal policy for $x_0 = D$ appears to be better as conjectured, by the fact that out of 40 trials with $x_0 = D$, the optimal and suboptimal policies gave the same x_N^2 in 21 trials, but in similar 40 trials with $x_0 = -D$, the optimal and suboptimal policies did not give the same x_N^2 in any case.

Although these results are by no means conclusive, they tend to support the conjecture that the one-stage suboptimal policy is a fairly good approximation to the optimal policy for the control system under consideration.

The adoption of this suboptimal policy simplifies the control problem considerably.

The approximate analysis of the control system with the one-stage policy will next be

* N is taken to be 12. For N > 12, the system stays pinned to either x = D or x = -D because of the particular boundary condition of the system until the number of remaining stages, n, becomes less than 12 or 11.

presented.

The Control System Behavior With One-Stage Policy

The state variable under the suboptimal policy is given by

$$x_{n+1} = \alpha x_n + r_n + v_n \qquad (3')$$

where $v_n = -m \cdot \mathrm{sgn}\left(\alpha x_n + (2p-1)b\right)$

$$r_n \begin{cases} = +b & \text{with prob. } p \\ = -b & \text{with prob. } 1-p \end{cases}$$

$0 < b < m, \quad 0 < \alpha < 1.$

Let us note that if $x_n > 0$ then $x_{n+1} < x_n$ and if $x_n < x_c$, then $x_{n+1} > x_n$ where $x_c = -(2p-1) \cdot b/\alpha$, because $(m-b) > 0$ by assumption. Namely, if the system deviation from the origin (the position of the equilibrium in the absence of the random disturbance r) is positive, the deviation at the next stage will invariably be smaller than the present one. If the present deviation is less than x_c, then the system will move towards the origin at the next stage.

For $x_c < x_n < 0$, $v_n = -m$

and $x_{n+1} > \alpha x_c - b - m = -(m + 2pb)$.

That is to say, once x_n becomes greater than $x_L = \mathrm{Min}\left[-(m + 2pb), -(2p-1)\, b/\alpha\right]$, then x_k cannot become smaller than x_L, for all $k \geq n$, i.e., x_L is the lower bound. Similarly, one sees that

$$x_u = m + 2b(1-p)$$

provides an upper bound. That is to say, if $x_n \leq x_u$, then $x_k \leq x_u$ for all $k \geq n$.

Figure (3) shows the typical system behavior using the one-stage suboptimal policy.

One noticeable feature of the system behavior is that x_n approaches the origin for a few stages consecutively and then x_n jumps away from the origin and from there it again approaches the origin for the next few stages.

Let us now investigate the number of stages before a jump occurs. This serves to show how the jerkiness of the system depends on system parameters, α, b, m and p.

Suppose that $x_L = -(m + 2pb)$, and $x_o = x_L$.

Let k be the first time that $x_k > x_c$ is realized.

The largest k will be realized when r_n is equal to - b for all k stages and the smallest k will be given when $r_n = +b$ occurs consecutively.

In the former

$$x_k = -\alpha^k x_o + \frac{(m-b)(1-\alpha^k)}{1-\alpha} \geq -\frac{(2p-1)b}{\alpha} \qquad (13)$$

In the latter

$$x_k = -\alpha^k x_o + \frac{(m+b)(1-\alpha^k)}{1-\alpha} \geq -\frac{(2p-1)b}{\alpha} \qquad (14)$$

From (13)

$$k \leq \frac{1n\left[\frac{m-b}{1-\alpha} + \frac{(2p-1)b}{\alpha}\right] - 1n\left[m+2pb+\frac{m-b}{1-\alpha}\right]}{1n\,[\alpha]} \qquad (15)$$

From (14)

$$k \leq \frac{1n\left[\frac{m+b}{1-\alpha} + \frac{(2p-1)b}{\alpha}\right] - 1n\left[m+2pb+\frac{m+b}{1-\alpha}\right]}{1n\,[\alpha]} \qquad (16)$$

With $\alpha = 7/8$, $p = 0.65$, $b = 1/16$ and $m = 9/128$, $k=5$ in (15) and $k=1$ in (16). $k = 5$ is realized with probability $(1-p)^5$ and $k=1$ is realized with probability p.

If $x_o = x_u$, then the similar argument gives

$$k \leq \frac{1n\left[\frac{m \mp b}{1-\alpha} - \frac{(2p-1)b}{\alpha}\right] - 1n\left[m+2b(1-p)+\frac{m \mp b}{1-\alpha}\right]}{1n\,[\alpha]} \qquad (17)$$

the maximum k is given with - sign and the minimum k is given with + sign in (17). The maximum k occurs with prob. p^k and the minimum k with prob. $(1-p)^k$.

Thus, after n becomes such that $x_L \leq x_n \leq x_u$ is realized, then the bounds on the number of consecutive stages, k, where the system approaches the origin monotomically are obtained.

Conclusions

It was shown that for a class of final value systems considered in this paper, $k_n(x, z)$ is concave in the parameter z which represents the a priori knowledge of the states of nature.

Policies derived from the functional values $z\, h_n(x;p_1) + (1-z)\, h_n(x;p_2)$ are shown to serve as initial suboptimal policy for adaptive control systems.

When the control variable is of the binary form $\pm m$, no explicit dependence of $h_n(x;p_i)$ and $v_n(x, p_i)$ on x is obtained.

However, one-stage suboptimal policy was shown to be a fairly good approximation to the optimal policy and was used to establish lower and upper bounds on the system deviations.

Acknowledgements

The author is indebted to Professor Gerald Estrin of the University of California, Los An-

geles, who first suggested the study of dynamic programming and was very helpful in numerous ways in the course of the investigation.

The author would also like to express his gratitude to Dr. R. Bellman, Dr. R. Kalaba of the RAND Corporation, and to Professor G. W. Brown of the University of California for their many helpful discussions.

The work reported here was supported in part by the Office of Naval Ordnance under contract Number Nonr 233(52) and 233(24). Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

References

1. Bellman, R. Dynamic Programming. Princeton University Press '57

2. Bellman, R. "Dynamic Programming and Stochastic Control Processes," Information and Control 1: 228-239 (Sept. 1958)

3. Aoki, M. Dynamic Programming and Numerical Experimentations as Applied to Adaptive Control Systems, Ph. D thesis, Department of Engineering, University of California, Los Angeles, Nov. 1959

4. Bellman, R. and R. Kalaba. Dynamic Programming and Adaptive Processes-Mathematical Foundations, P-1416. RAND Corp., Santa Monica, Calif., 1958

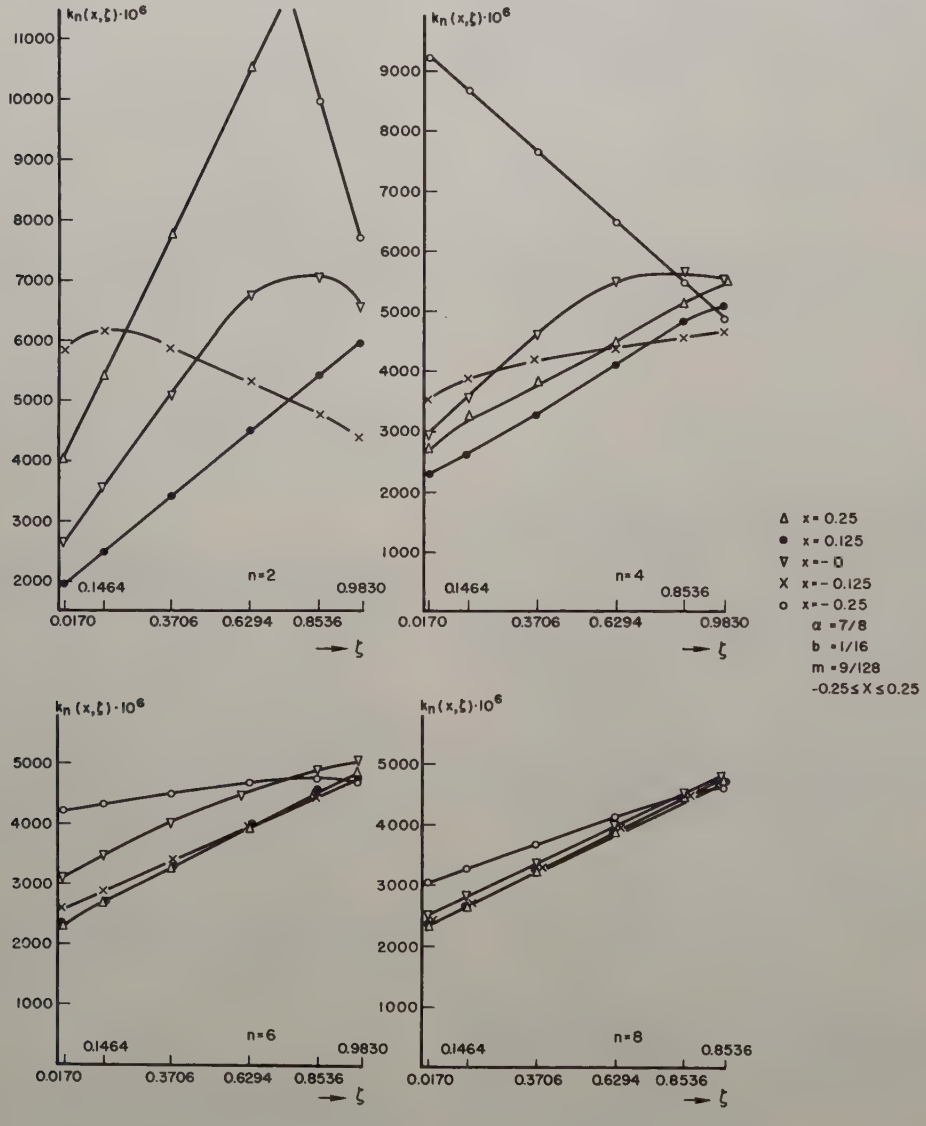


Fig. 1. The criterion function $k_n(x, \zeta)$ as the function of the a priori probability, ζ.

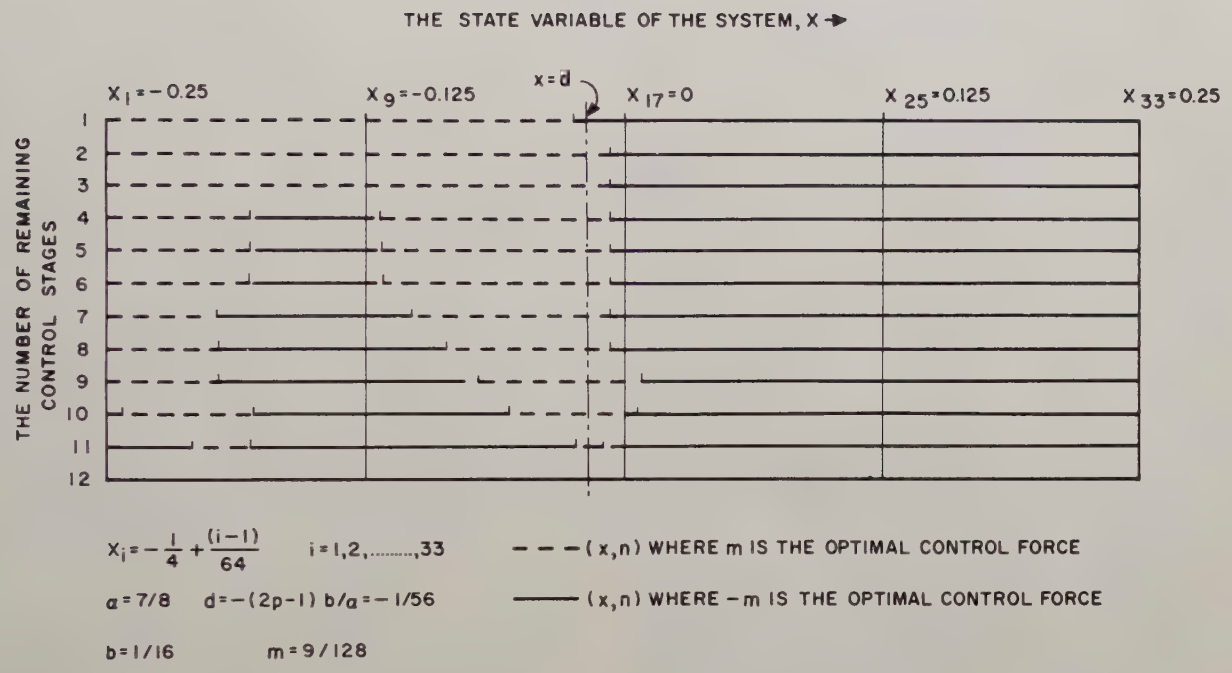


Fig. 2. Boundaries between +m control force and -m control force in the optimal and the sub-optimal policies at p=0.625.

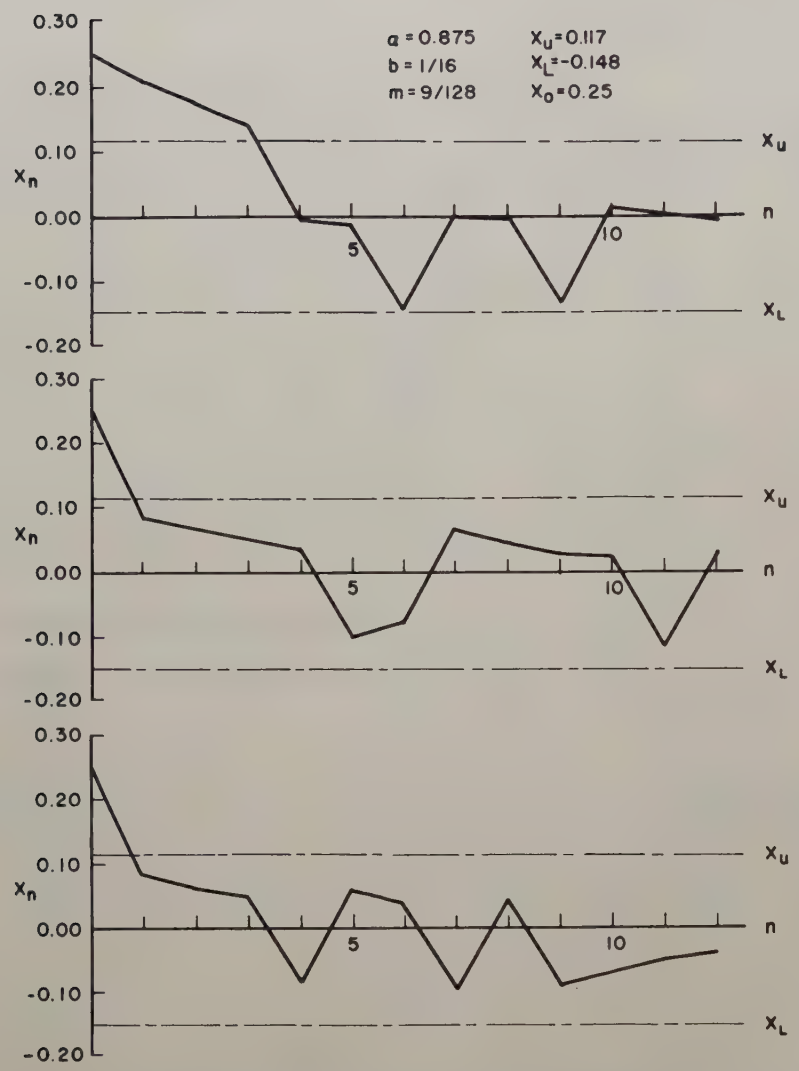


Fig. 3. Typical behavior of the state variable x_n, with the members of past control actions, n, at p=0.625.

A STUDY OF ASYNCHRONOUSLY EXCITED OSCILLATIONS IN NONLINEAR CONTROL SYSTEMS

Olle I. Elgerd
Department of Electrical Engineering
University of Florida
Gainesville, Florida

Summary

This study was initiated as an effort to explain certain oscillatory phenomena observed in an aircraft and weapon control system operated in a particular mode. The oscillations in question were of the undamped limit-cycle type and their presence could very clearly be correlated with the degree of noise corruption of the signals.

It is demonstrated in this paper that certain types of nonlinear systems although being inherently stable may be driven into oscillatory modes not only by random signals but also by any high frequency periodic or nonperiodic signal possessing a certain energy content. It is of interest to note that the tendency for hunting and also the hunt frequency usually are completely independent of the frequency of the excitation signal, i.e., the phenomenon is of asynchronous nature.

A general theory, the validity of which has been tested by both analog and digital means, is presented and utilized to demonstrate how this phenomenon may be predicted from information on circuit data.

Introduction

Although the phenomena to be discussed in this paper can only occur in systems incorporating certain types of nonlinear components, it should nevertheless prove enlightening to take a quick look at some of the characteristic features of linear systems. This will be done in order to demonstrate the uniqueness of the phenomenon and also in the hope that it may be possible to employ some of the well-established linear analysis techniques in explaining it.

A system is defined as linear if its dynamical behavior can be described by a system of linear differential equations, partial or ordinary. From the analyst's point of view, a linear system is quite attractive because of the relative ease with which its behavior can be predicted. The control engineer in particular is appreciative of the following two characteristics:

1. The stability of a linear system is uniquely determined from the parameters of the system only.

2. In response studies the super-position rule can be employed.

These two features, probably to a higher degree than any others, put the linear system in sharp contrast with the nonlinear one. The first one of the above rules states that a linear system cannot be driven into periodic or growing oscillations by any combination of external input signals of finite duration. One may, in contrast, quite simply demonstrate a multitude of nonlinear systems which, for instance, are stable for low signal levels but go into limit cycles (hunting) should they be excited by high level inputs.

The second rule points out a very important "nonmixing" characteristic of a linear (stable) system which indeed constitutes the basis for all the well-known frequency response techniques in use. A linear system will at its output terminals contain only those frequencies it has received at its input. A nonlinear system, on the other hand, never displays such a frequency discrimination. It may accept a single or a few frequencies at its input and after a complicated mixing process deliver an infinite, but discrete, spectrum at its output. The new frequencies created in this process are referred to as a modulation product. It will prove advantageous in the following to consider the nonlinear components as modulators.

It has been known for considerable time that injection into a control system of a high frequency dither signal will in certain instances upgrade the performance of the system. The frequency of this signal is usually chosen considerably above the cutoff of the system so that its presence cannot be detected in the output. The "dynamical lubrication" provided by such a signal was first pointed out by MacColl[1] who analyzed its effect in an on-off type servo. It is today quite common to diminish the effect of static friction by use of a superimposed high frequency signal. (This "lubrication effect" is explained in a simple fashion by the theory later to be developed in this report.)

MacColl also pointed out how, by inserting a high frequency signal into the on-off servo mentioned above, one could improve its sensitivity for small signals. Oldenburger had discovered experimentally that a high frequency signal superimposed upon the input signal of a

nonlinear governor control system tended to improve its stability. Indeed it was possible to completely quench self-oscillations under certain conditions. He reported his findings in two separate papers[2,3], in the latest of which he attempts to analyze a particular single-loop system with a well-defined nonlinearity in the forward loop. This method of stabilizing nonlinear systems subject to limit cycle oscillations is known under the name of "Signal Stabilization".

The topic of the present investigation is to study the opposite problem: Is it possible to excite an inherently stable system by a high frequency noise signal in such a manner that the system ceases to be stable? To the author's knowledge, this problem has never concerned control people. As a matter of fact, an extensive search of the literature on nonlinear phenomena reveals very little work in this area. It should be made clear at this point that the phenomena discussed here ("Signal Stabilization" included) differ essentially from the well-known phenomenon of so-called nonlinear or "subharmonic resonance" in which the existence of a rational ratio between system frequency and frequency of excitation always must exist.

A very extensive review of subharmonic resonance phenomena is given by Ku in his recent book[4]. Actually when one realizes the very large interest shown in subharmonic resonance by so many investigators throughout the years, one is surprised to find only one reference discussing a phenomenon quite similar to the one to be taken up for study in the present report. Minorsky[5] made, in 1955, a study of a nonlinear system obeying Van der Pol's differential equation. By applying an external harmonic excitation to his system he found what he describes as an "asynchronous action". It manifests itself in the following two ways:

1. A system subject to limit cycles may have its oscillations quenched by the externally applied frequency.

2. A system may be brought into limit cycle oscillations by the application of the external frequency.

The word "asynchronous" implies that there need not exist any rational ratio between the external frequency and the limit-cycle frequency. This particular feature brings Minorsky's findings in close relation to the ones to be reported in this paper. There are, however, important differences present. Generally the above-mentioned asynchronous "quenching" and "excitation" actions cannot occur in the same system; Minorsky shows that one has to make a distinction between the so-called "soft" and "hard" systems. Furthermore, he makes no mention that the system he studied will be neither excited nor quenched by a stationary random signal.

Objective of Present Investigation

After having demonstrated a simple control system that can be excited in an asynchronous manner, a general explanation of the phenomenon will be given. To understand a phenomenon is to have in one's possession a theory with a proven capability of prediction. In order to extend the ability of prediction the author has not considered phase-plane techniques which, although used by Minorsky in his above-mentioned work, have a very limited range of applicability. Instead, an analysis method has been developed which permits a study of systems with considerably higher degrees of complexity.

As was mentioned earlier, this research was actually initiated as an effort to rid a particular system of noise-excited oscillations. For this reason the report discusses methods of obtaining most effective suppression of undesired oscillations of asynchronous nature.

The author feels that there might be some possibility of utilizing asynchronous action for improvements of system response. These thoughts are expounded in this paper.

Theory

Consider the simple control system depicted in Figure 1. Its forward loop incorporates a linear (2) and a nonlinear (1) component. The linear component has a transfer function:

$$G_2(s) = \frac{K}{s(1+s)^2} \qquad (1)$$

$$(K = \text{real})$$

The nonlinear component may represent an amplifier having a gain-versus-signal-level characteristic as shown in Figure 2.

Assuming that the system is operated in such a way that the error e(t) will not exceed e_o, i.e., the breakpoint of the gain curve, the overall gain will be:

$$K_{total} = K \cdot 1 = K$$

With the above limitation put on the signal level it is clear that the overall system can be analyzed as a linear system. In Figure 3 are drawn the Nyquist plots for three different values of the gain. The system is obviously stable for $K < 2$. The gain is now set equal to unity corresponding to a gain margin of 100 percent and the system response is studied on an

analog computer. The nonlinearity is simulated by means of diodes.

Figure 4 depicts a series of seven computer records. The four simultaneous recordings show in this order:

$r(t)$ = input

$c(t)$ = response

$e(t)$ = error signal = $r(t) - c(t)$

$e_1(t)$ = modified error signal (defined in Figure 1)

Graph 4A indicates the system response for a step input and shows clearly the stable nature of this system. Without changing the amplitude of the step input, the response is then studied with a high frequency dither signal superimposed upon the input. Graphs 4B, C, D and E depict the change in response for increasing amplitude of this dither signal. Two interesting observations are made from these graphs:

1. The dither frequency is so high that no trace of it can be found in the system output $c(t)$ because of the filtering effect of G_2.

2. When the dither signal reaches such an amplitude that its added effect on $r(t)$ brings the nonlinear element beyond its breakpoint, (can be seen on the e_1 recording) the system turns more and more oscillatory.

Should the dither signal be given large enough amplitudes, the system will be driven into instability as shown in graph 4F. This graph indicates actually that instability occurs even without a signal input, i.e., it behaves in this respect as an unstable linear system where oscillations start to build up from a quiescent state. Should the dither signal be removed after the system has worked itself up to a certain amplitude of oscillations, the oscillations would either die out or continue to grow. The latter would happen if the signal were removed so late that the (low frequency) oscillations themselves would be large enough to bring the error signal sufficiently beyond the breakpoint.

Graph 4G finally proves what was stated earlier, namely that oscillations can be excited also by a noise signal.

The example just demonstrated referred to a particular system exhibiting a particular nonlinearity. In attempting to explain this phenomenon, we should search for a method enabling us to preserve a high degree of generality. In spite of the difficulty to generalize nonlinear system behaviors, it has been possible to devise a technique which permits prediction of the system response in the presence of high frequency dither. The method is an extension of the so-called describing function technique.

The describing function method has been proved quite useful in study of nonlinear systems. With some modification, this method can also be applied to gain an understanding of the present phenomena. The keystone of the describing function analysis is an assumption, the value of which is usually guaranteed by the frequency discriminating characteristic of a normal type control system. The method is usually capable of predicting not only the possible existence of sustained oscillations but also the corresponding (approximate) amplitudes and frequency. It does not predict the waveshape and is of little value in determining transient response.

The "black box" N in Figure 5 represents a nonlinearity the only restriction on which is that it must be non-time varying. The effect of this restriction is that any sinusoidal input will result in a periodic output the period time of which equals the period time of the input. We may write the output as a simple Fourier series:

$$y(t) = \sum_{n=o}^{\infty} B_n \sin(n\omega t + \Phi_n) \qquad (2)$$

for an input

$$x(t) = A \cdot \sin \omega t$$

By focusing attention upon only the fundamental frequency component in the output one defines a "describing function" N,

$$N = \frac{B_1}{A} \cdot e^{j\Phi_1} \qquad (3)$$

which function indicates how the signal changes amplitude and phase by passing the "black box". From then on, linear analysis techniques are employed[6]. One normally has to take into account the fact that N changes with both A and ω.

After this sketchy review of the describing function technique, we return to our actual problem. Consider the "black box" in Figure 5 again. Let the input now consist of a sum of two separate frequencies:

$$x(t) = A \sin \omega t + A_o \sin \omega_o t \qquad (4)$$

The output can now be described by a double Fourier series[7].

$$y(t) = \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} B_{m,n} \cdot \sin \left[(m\omega_o + n\omega)t + \Phi_{m,n} \right] \tag{5}$$

This statement is not self-evident and requires some explanation. A simple heuristic proof is given as follows: the nonlinear relationship of the "black box" may be expressed as an infinite power series:

$$y = \sum_{i=0}^{\infty} a_i x^i \tag{6}$$

This is sometimes referred to as "curve fitting". Usually very few terms are sufficient to accurately describe the nonlinearity. By combining Equations (4) and (6) one may obtain the output as a function of the input frequencies:

$$y(t) = \sum_{i=0}^{\infty} a_i \left(A \sin \omega t + A_o \sin \omega_o t \right)^i \tag{7}$$

For increasing i-values one obtains (in addition to a constant) terms of the following types:

$\sin \omega t$

$\sin \omega_o t$

$\sin^2 \omega t$

$\sin^2 \omega_o t$

$\sin \omega t \cdot \sin \omega_o t$

$\sin^3 \omega t$

$\sin^3 \omega_o t$

$\sin^2 \omega t \cdot \sin \omega_o t$

$\sin \omega t \sin^2 \omega_o t$

etcetera

In other words, the output contains (in addition to the d-c component) the following frequencies:

ω

ω_o

2ω

$2\omega_o$

$\omega \pm \omega_o$

3ω

$3\omega_o$

$\omega \pm 2\omega_o$

$\omega_o \pm 2\omega$

etcetera

This is indeed what formula 5 implies. The validity of the formula can be verified in the case where the nonlinearity is of memory type, for instance -- hysteresis.

The output spectrum is obviously considerably more dense than in the case of only one input frequency. Computation of the different components is associated with great numerical difficulty, even in the case of a simple nonlinearity. This discouraging fact will not, however, prevent us from drawing some general conclusions of great importance.

It should be remembered that in the case under study it is generally true that there is a considerable difference in magnitude between the input frequencies, i.e.:

$$\omega_o \gg \omega$$

Furthermore, it will be stated without proof that the coefficients $B_{m,n}$ in the series 5 are of diminishing magnitude with increasing values of m and n. (This situation is well-known in the case of a simple Fourier expansion.) Generally this means that the dominant frequency in the lower region of the output spectrum is ω. This very important conclusion can be demonstrated for a representative numerical case. Assume:

ω = 10 radians per second

ω_o = 105 radians per second.

The output spectrum, as obtained from formula 5, contains in increasing order of magnitude the following angular frequencies (d-c component neglected) :

5 rps (corresponding to m = 1, n = 10)

10 rps (corresponding to m = 0, n = 1)

15 rps (corresponding to m = 1, n = 11).

The first and third components, corresponding to the relative high n values of 10 and 11, respectively, may be neglected, and consequently the dominant lower component in the output spectrum is the one having a frequency equaling the lowest input frequency = 10 rps. Corresponding amplitude is $B_{o,1}$.

At this point we shall now make the same basic assumption as one always makes in order to use the describing function technique: due to the filtering effect of cascaded components in the overall system, all relatively high frequency components will be neglected.

This assumption combined with previous findings relating relative magnitudes of spectral components gives as final result that the only output component of any importance will be $B_{o,1}$.

It should be clear at this point that it now is possible to define a "describing function" in an analogous fashion as was done in Equation (3):

$$N = \frac{B_{o,1}}{A} e^{j\Phi_{o,1}} \qquad (8)$$

There is, of course, a fundamental difference between the two describing functions defined by Equations (3) and (8), respectively. Whereas the function defined by Equation (3) depends only on ω and A, the new describing function will depend not only on those two variables but also upon A_o and ω_o, i.e., the amplitude and frequency of the dither signal. For the purpose of identification the function defined by formula 8 will be referred to as "modified describing function". For most types of nonlinearities, one example being the one shown in Figure 2, the dependence on the dither amplitude will be pronounced indeed. The dither frequency, on the other hand, usually has a nonsignificant effect.

The last two statements are very important. In fact, they provide the very clue for a proper understanding of not only the asynchronous excitation phenomenon, but also the phenomenon of signal stabilization. The presence of the dither signal controls in effect the gain and phase characteristics of the nonlinear element, as felt by the low frequency signal. The gain and/or the phase of the element may be controlled by varying the amplitude but not the frequency of the dither signal. This influence of the dither amplitude on the low frequency gain and/or phase is indeed most interesting. Borrowing an analogy from the field of chemistry, one might think of the dither amplitude as a catalyst, the presence of which is necessary for a certain process to take place. The independence of the dither frequency explains immediately the asynchronous characteristics of the phenomenon. The only criterion is that this frequency is high enough.

Numerical Evaluation of the Modified Describing Function

So far the discussion has been entirely of a qualitative nature. We shall now proceed with a presentation of a method aimed at obtaining numerical information in any particular case of interest. From what was said in a previous section, the effect of dither injection can be uniquely predicted on basis of the magnitude and phase of the modified describing function N in formula 8, knowledge of which presupposes that the amplitude $B_{o,1}$ and the phaselag angle $\Phi_{o,1}$ be known.

The problem of calculating the amplitudes $B_{m,n}$ in the double Fourier series 5 should be expected to turn out to be considerably more difficult than the analogous problem relating to a simple Fourier series. The reader is reminded that the amplitudes in the latter case can be calculated from a simple integral extended over the period-time of the function being Fourier analyzed. In the present case of a double Fourier series, one makes the discouraging discovery that the amplitude $B_{m,n}$ must be evaluated from double integrals. It would seem at a first glance, then, that this would mean an effective stop to any further analysis attempt. Fortunately, this conclusion is not correct. A considerable simplification can be achieved by once more making use of the assumption that $\omega_o \gg \omega$. In this case the problem of determining $B_{o,1}$ will be reduced to solving a simple integral. This situation will be demonstrated with a simple example. Consider again the system depicted in Figures 1 and 2 which was discussed in the first part of the previous section. The recordings shown in Figure 6 indicate the input to and the output from the nonlinear element of this system. In this particular case we have

$$\omega_o \approx 33\omega$$

The output graph clearly indicates a periodicity corresponding to the lowest frequency in the input, i.e., ω rps. However, in a mathematical sense, no periodicity is present; the existence of such a periodicity would require that ω_o were an even multiple of ω. In reality, the output contains frequencies below ω which one realizes immediately from formula 5; one can quite easily find integers m and n satisfying the equation.

$$m\,\omega_o + n\omega < \omega$$

These lower frequencies are indeed present in the output wave of Figure 6 but are of such a small magnitude that they cannot be detected. (The reader is here reminded of the earlier discussion where this situation was explained.)

It seems intuitively justifiable to attempt to calculate $B_{0,1}$ under the assumption that ω is the lowest frequency and thus employ the simple formulas of simple Fourier analysis. In other words, the output wave of Figure 6 is assumed periodic with a period-time of

$$T = \frac{\omega}{2\pi}$$

We obtain

$$B_{0,1} = \frac{2}{T} \int_0^T y(t) \cdot \sin \omega t \, dt \qquad (9)$$

It can be shown that in general this formula is quite exact assuming that $\omega_0 \gg \omega$. Although formula 9 represents the ultimate in simplicity that we can expect to achieve, there might yet be a big numerical step left to take. At this point a digital computer comes in very handy. The curves in Figure 7 show in normalized units the modified describing function of the above-mentioned nonlinearity. The absolute value of N is plotted versus dither amplitude (A_0) and signal amplitude (A). The phase of N is, of course, zero for this type of nonlinearity. A few words should be said about the numerical integration necessary in this and similar cases. The factor y(t) appearing in the integrand will even for relatively simple types of nonlinearities defy a simple analytical description. This fact is readily appreciated after a look at the output waveform in Figure 6. Each time breakpoint of the nonlinearity is passed, and this occurs a large number of times throughout the time interval T, a discontinuous shift takes place in the derivative of the output. If the integral 9 were to be solved analytically, one would have to first pinpoint all those intervals and then perform the overall integration by subdividing the integral into a large number of part integrations. This difficulty is bypassed if a digital computer is utilized. In preparing the integral for a computer program, it is first written as a sum:

$$B_{0,1} = \frac{2}{T} \int_0^T y(t) \sin \omega t \, dt \approx$$

$$\approx \frac{2}{n} \sum_{i=0}^{n}{}' \; y\left(i \frac{T}{n}\right) \sin \omega \left(i \frac{T}{n}\right) \qquad (10)$$

n is the number of subdivisions of the interval 0-T. The choice of n is largely dependent on the ratio ω_0/ω, because one must make the intervals sufficiently small to accurately account for the change in the most rapidly changing term, i.e., be one due to sin ω_0. The program is then worked out in such a way that the computer in sequence computes and adds up terms for increasing values of i. In so doing, it should make a logical choice of proper value of

$$y\left(i \frac{T}{n}\right)$$

in accordance with:

If $x(t) \leq |e_0|$ choose $y(t) = x(t)$

If $x(t) > e_0$ choose $y(t) =$

$= e_0 + (x(t) - e_0) \tan \alpha_2$;

If $x(t) < -e_0$ choose $y(t) =$

$= -\left[e_0 + (x(t) - e_0) \tan \alpha_2\right]$

Because of the inherent capability of the computer to make such a choice a numerical evaluation of $B_{0,1}$ will be quite simple. The curves in Figure 7 represent 35 different computed values of $B_{0,1}$. In each case a subdivision into 7200 parts was used. Total computing time was 10 minutes on an IBM 709.

It is interesting to note that the performance predicted from this data was in perfect agreement with results obtained on the analog computer.

Effect of Various Types of Nonlinearities

In the example just studied the presence of dither caused an increase of system gain with a resulting loss of system stability. Other types of nonlinearities might give rise to different effects. The saturation type of nonlinearity is quite interesting from several points of view. From a historical point of view, this type of nonlinearity, depicted in Figure 8, was present in a system where signal stabilization first was used[2]. The graphs in the same figure indicate the modified describing function calculated for this type of an element. As might be intuitively felt, the gain decreases with both signal level and dither amplitude. This is, of course, the explanation why Oldenburger was able to detect a stabilizing effect in a system incorporating this type of nonlinearity.

In order to further test the validity of the theory presented in this paper, a particular system was simulated which contained the saturation type nonlinearity. The system is shown in Figure 9 and its response with and without dither injections in Figure 10. Instead of obtaining a stabilizing effect as might have been expected on basis of Dr. Oldenburger's findings, this system obviously goes unstable

in the presence of dither. Actually, this is exactly what can be predicted on closer study. The linear part of the system is namely dynamically conditionally stable, i.e., it will actually be driven into an unstability region by decreasing the gain. This fact is easily verified by a Nyquist plot. This example brings out quite sharply the important point that in analyzing the effect of a nonlinearity, one definitely must take into proper account the linear portions of the system.

Both of the above-mentioned types of nonlinearities have real describing functions. It is interesting to study the effect of dither in a case where the nonlinearity is of memory type, i.e., the present output is a function of the past history of the input.

Backlash is a commonly encountered representative for this type of nonlinearity. Dither will cause a very interesting change in the modified describing function for such an element. Figure 11 contains a number of plots computed on the IBM 709. As in earlier diagrams of the same type all symbols are defined in the Figure. The graphs reveal that dither will increase the gain. In addition, an effect can be noted in the phase. The backlash introduces normally a phaselag, the magnitude of which increases with the magnitude of the backlash (2C in Figure 11). The dither will decrease this phaselag; in the case where the dither amplitude equals half the magnitude of the backlash, the phaselag will be zero.

This dual effect will quite clearly affect the system response. Figure 12 shows a specific case. The system is identical with the one in Figure 1 where the backlash has been introduced in the nonlinear "black box". In this case, the dither will obviously improve the stability, the reason being that the stabilizing decrease of phaselag will dominate over the de-stabilizing increase of gain. It would be easy to think up systems where the opposite effect would be detectable. In Figure 12C is shown the effect on the same system when the dither is made up by noise.

Backlash represents the most complicated case for which the modified describing function has been computed in this report. It may be of interest, therefore, to include the computer program (FORTRAN) to give a general idea of degree of simplicity (see Figure 13).

Effect of Random Type Dither

In the previous theoretical treatment, attention has been limited to dither composed of only one harmonic component having a frequency of considerably greater magnitude than the cutoff frequency of the control system. It was shown that the modulation products could be evaluated albeit the numerical effect could become quite strenuous. If several high frequency components are present, the overall effect on the output from the nonlinear element will be that more frequencies will be present, i.e., the output spectrum will become denser. Due to the fact that superposition cannot be employed, the problem of numerically evaluating the modulation product of interest becomes increasingly difficult. For example, should we add just one more component to the dither in the previously treated, relatively simple nonlinear cases, we would find that a numerical computation of the modified describing function would be an almost impossible task, even using highspeed digital computers.

Random noise, from a frequency viewpoint, consists of a continuous spectrum. Even if the frequency distribution were known, an attempt to find analytically the describing function using methods presented in this paper are doomed to meet with failure.

A different approach based on statistical considerations was tested but failed to give usable results. By "usable" are meant formulas which would render numerical results with a reasonable effort. It is the author's opinion that this general problem is of such a nature that it defies practical computation.

From a practical point of view it may be of more value to report certain semi-empirical observations that were made during the progress of the research. A large number of analog computer runs were made in order to test the theory. The effect of noise corruption was studied for every configuration that was tested. The following important observations were made:

1. Whatever the system behavior turned out to be in the presence of a single harmonic dither signal it remained identical for any type of dither waveshape including random noise.

2. The energy content of the dither signal (as measured by its RMS value) rather than its waveshape seemed to be determining for overall system behavior.

It is quite simple to prove by inductive reasoning that the first observation has general validity. If, for instance, a single harmonic dither signal can be shown to have a certain effect of the modified describing function, this effect will be increased by addition of, say, one or more high frequency components.

It is obviously of great value to be able to predict in a qualitative way what will happen in the presence of a noise signal. If we can postulate the generality of the second observation, it is also possible to make quantitative predictions based on a knowledge of signal energy content only. The simplicity of this

rule is, by the way, appealing from an intuitive standpoint. The rule states that one can calculate the effect on the modified describing function of a single harmonic dither signal, making full use of the techniques of this paper, and then assume that any high frequency signal possessing the same energy (RMS value) will cause the same change in the describing function. Of all possible cases tested on the analog computer, this rule gave results that were correct within ± 5 percent.

Filtering Considerations

When the asynchronous action degrades the performance of a system and it therefore proves desirable to rid the system of it as completely as possible, the best method to use is obviously the one that removes the cause, i.e., filters out the dither from the signal. The obvious conclusion that can be drawn from this study is that such a filter, in order to be of maximum efficiency, should be placed before the nonlinear element. If the filter were placed after this element, it could effectively reduce the high frequency portion of the signal, but too late to prevent its effect on the low frequency gain and phase. The fundamental rule for filtering would therefore be: reduce the dither content to a minimum before it reaches the nonlinear element. How effective a filter to chose can be determined from diagrams of the type shown in Figures 7, 8, and 11.

Asynchronous Action as a Useful Design Tool

Signal stabilization has been earlier mentioned as an example where asynchronous action manifests itself in a beneficial way. It can also be used to obtain an artificial increase in system sensitivity in cases where, for instance, "dead-zone" of backlash result in insensitive regions for small signal levels. When the phenomenon of asynchronous action will be better understood, and this paper makes a claim to furthering this knowledge considerably, it is entirely possible that new uses can be envisioned. One feature of asynchronous action is the possibility of obtaining a variable gain simply by varying the dither amplitude. One might expect possible uses of this characteristic in design of adaptive controllers for instance. It seems worthwhile also to investigate compensation networks where one utilizes the special features of nonlinear networks as, for instance, recently reported by Mishkin and Truxal[8] but combined with the additional possibilities of asynchronous action.

References

1. MacColl, LeRoy A., Fundamental Theory of Servomechanisms, New York, D. Van Nostrand Co., Inc., 1945, pp. 78-87.

2. Oldenburger, Rufus, "Signal Stabilization of a Control System," ASME Transactions, Vol. 79, 1957, pp. 1869-1872.

3. Oldenburger, Rufus, Lin., C.C., "Signal Stabilization of a Control System," AIEE Transactions, (Applications and Industry), May 1959, pp. 96-100.

4. Ku, Y.H., Analysis and Control of Nonlinear Systems, New York, The Ronald Press Co., 1958, pp. 175-190.

5. Minorsky, N., "On Asynchronous Action," Journal of the Franklin Institute, Vol. 259, 1955, pp. 209-219.

6. Truxal, John G., Control System Synthesis, New York, McGraw-Hill Book Co., 1955, pp. 559-612.

7. Bennett, W.R., "New Results in Calculation of Modulations Products," Bell System Technical Journal, Vol. 12, 1933, pp. 228-243.

8. Mishkin, E., Truxal, J.G., "Nonlinear Compensation Networks for Feedback Systems," IRE National Convention Record, Part 4, 1957, pp. 3-7.

Acknowledgement

The work reported in this paper was done at Hughes Aircraft Company, Culver City, California.

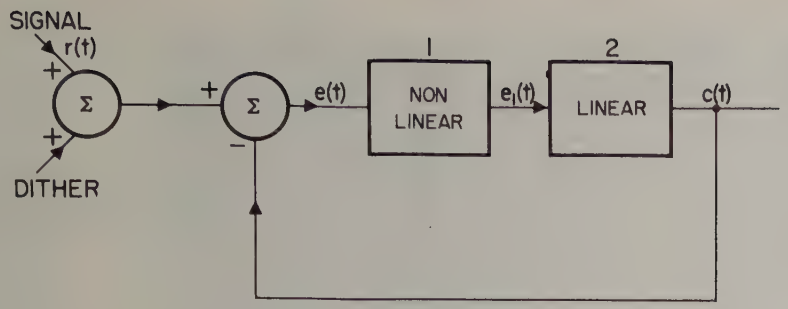


Fig. 1. System used for demostration of asynchronous action.

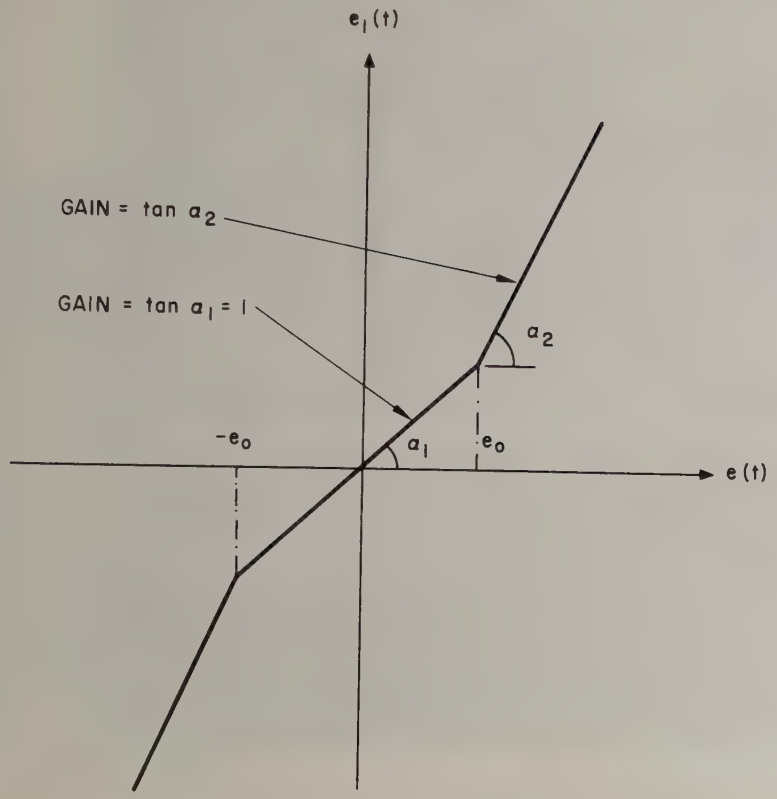


Fig. 2. Nonlinearity used in experiment outlined in text.

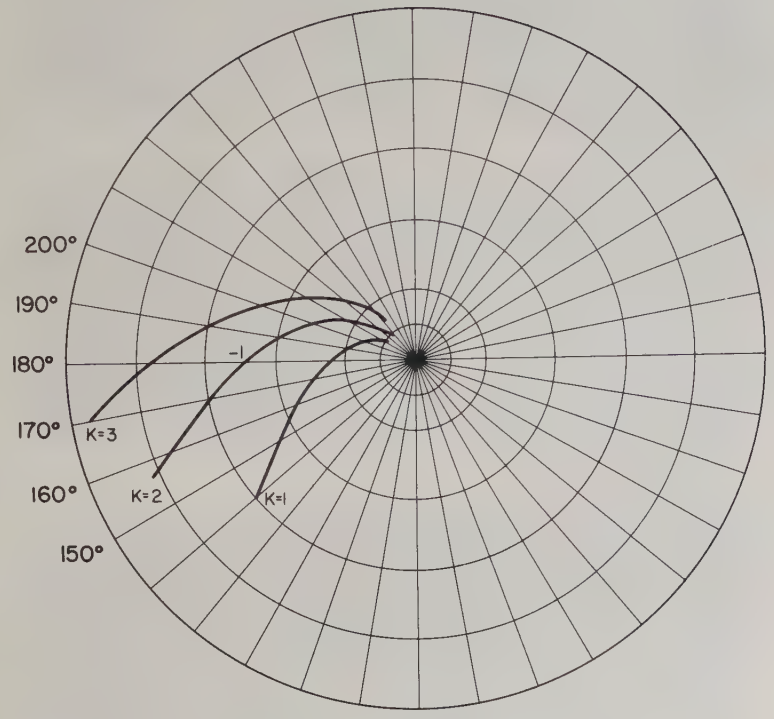


Fig. 3. Nyquist plots for system having a transfer function in accordance with equation (1).

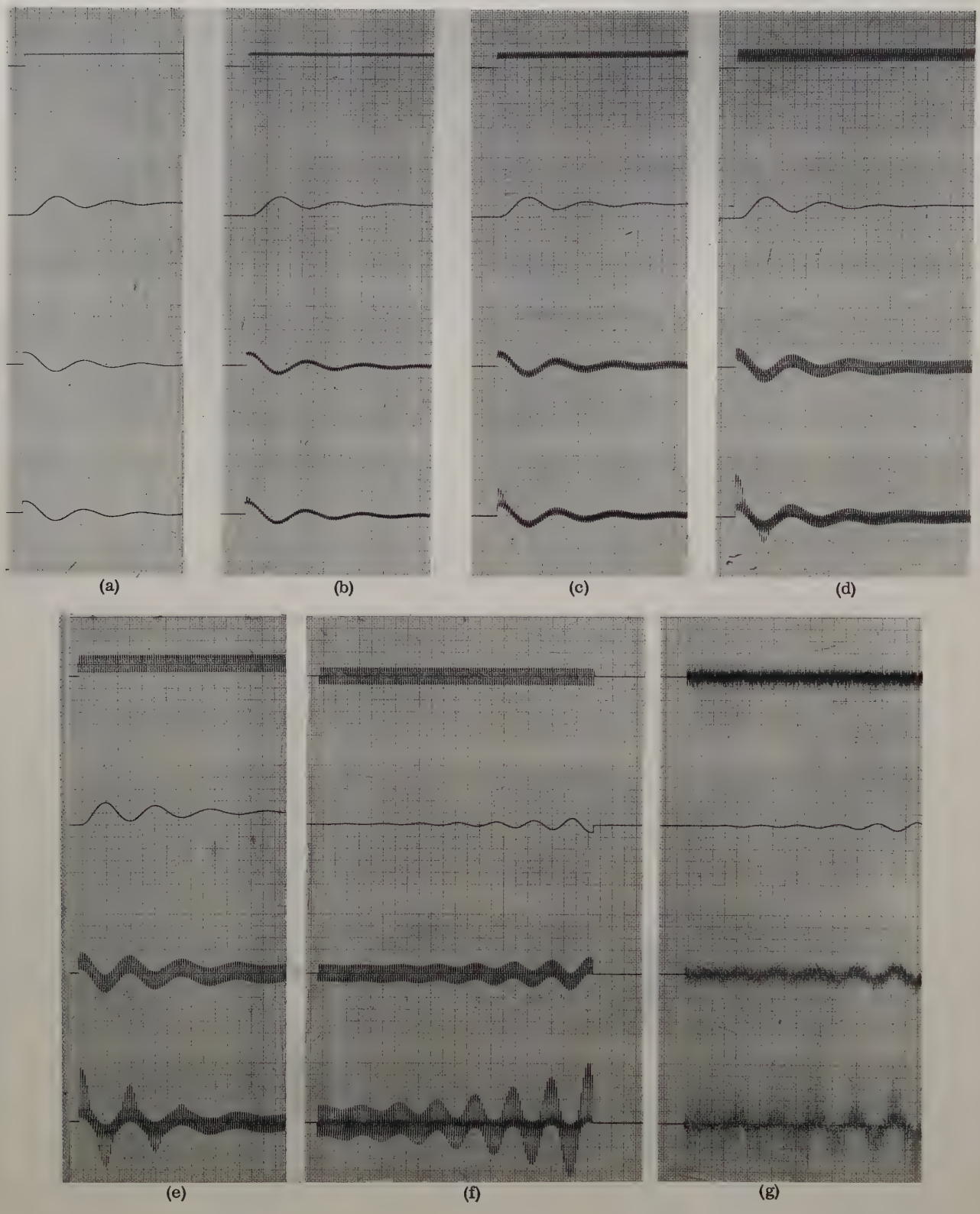


Fig. 4. Response of system shown in Fig. 1. (a)-(e) indicate change in response with increasing dither amplitude. Recordings (f) and (g) prove that oscillations actually will start to grow from a quiescent state, assuming the dither to be of sufficiently high level. The four different channels depict in order: r(t), c(t), e(t), and $e_1(t)$. This applies to all similar graphs of this paper.

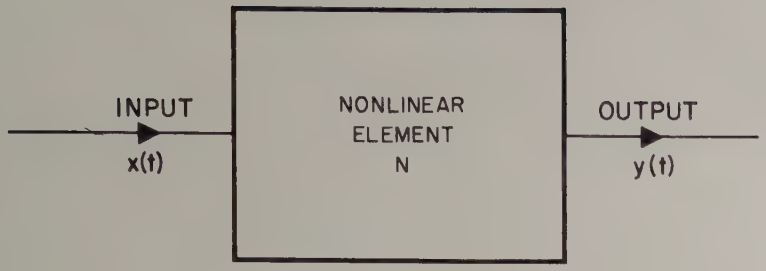


Fig. 5. Time-invariant nonlinearity.

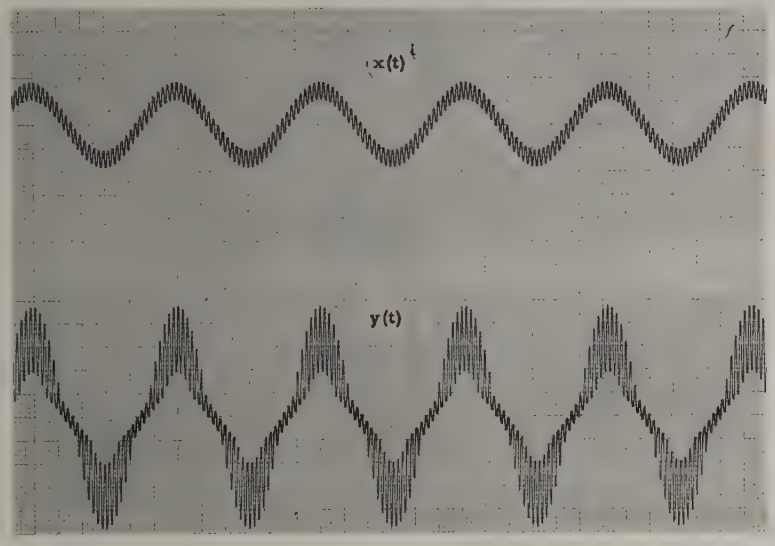


Fig. 6. Modulation of a two-frequency signal by the nonlinearity shown in Fig. 2.

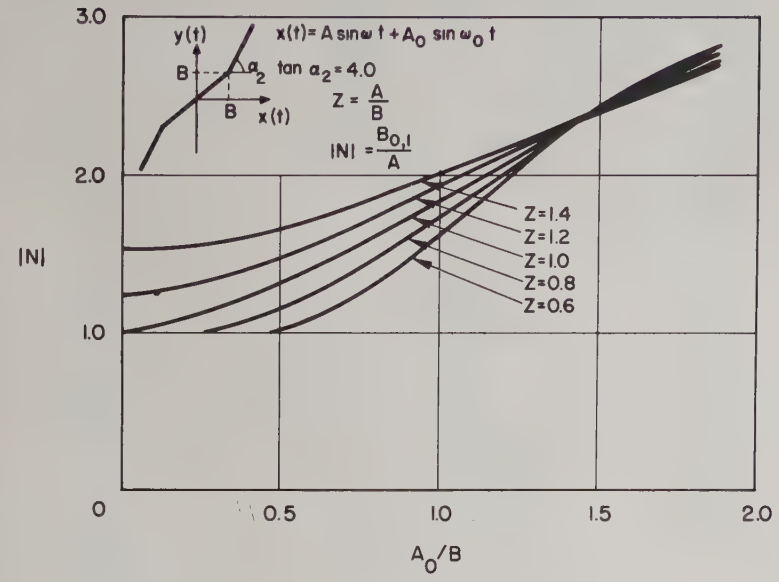


Fig. 7. The variation of the modified describing function with respect to dither amplitude and signal level for a nonlinearity depicted in the figure.

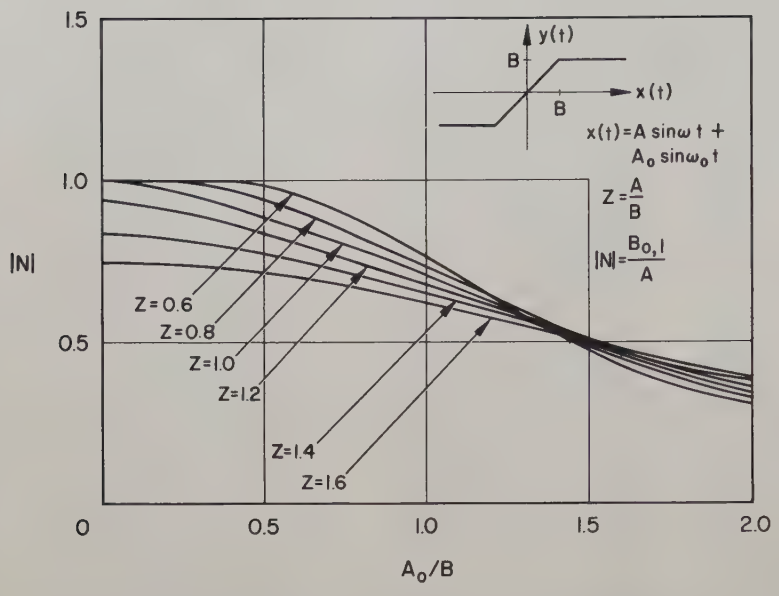


Fig. 8. The variation of the modified describing function with respect to dither amplitude and signal level for a saturation-type nonlinearity.

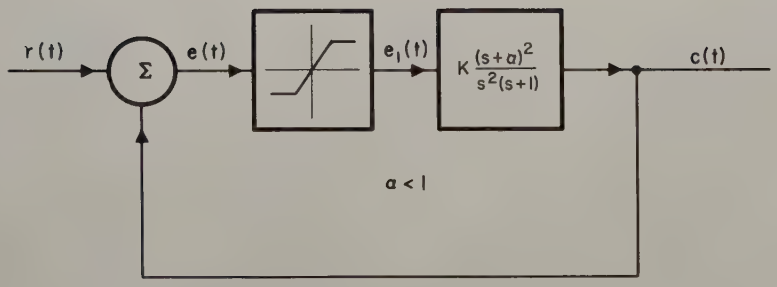


Fig. 9. A conditionally stable control system used for simulation.

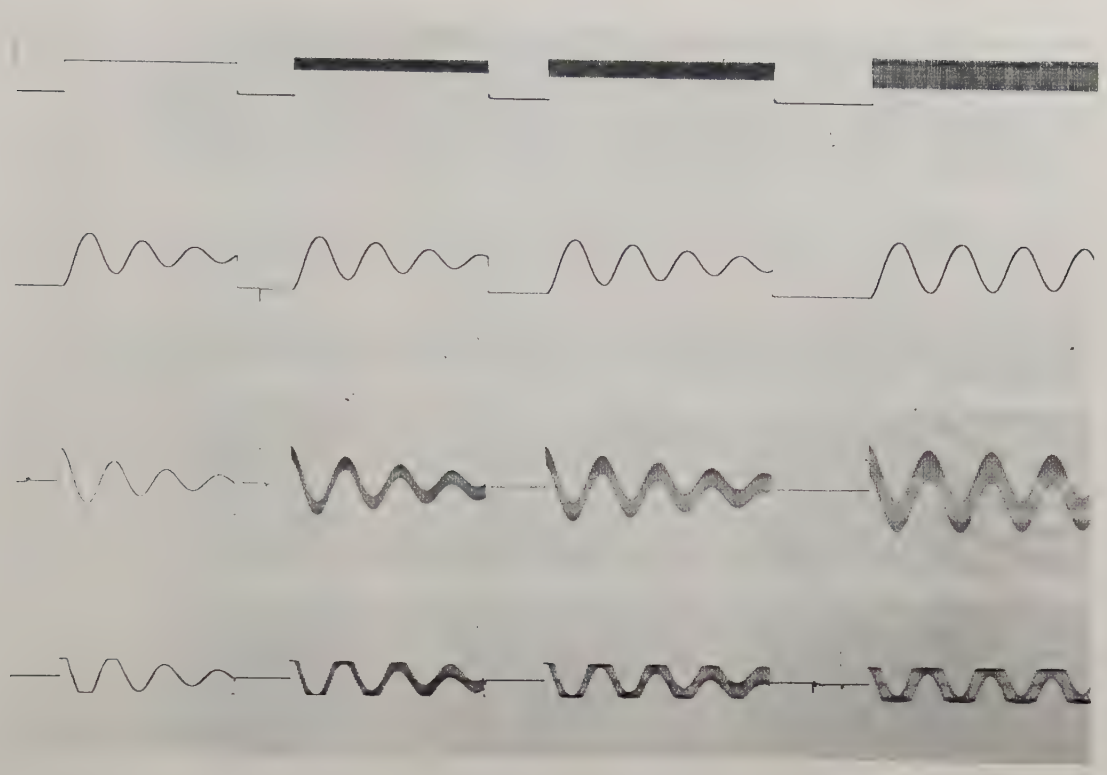


Fig. 10. Response of system shown in Fig. 9. One notices the decreasing amplitude of dither signal.

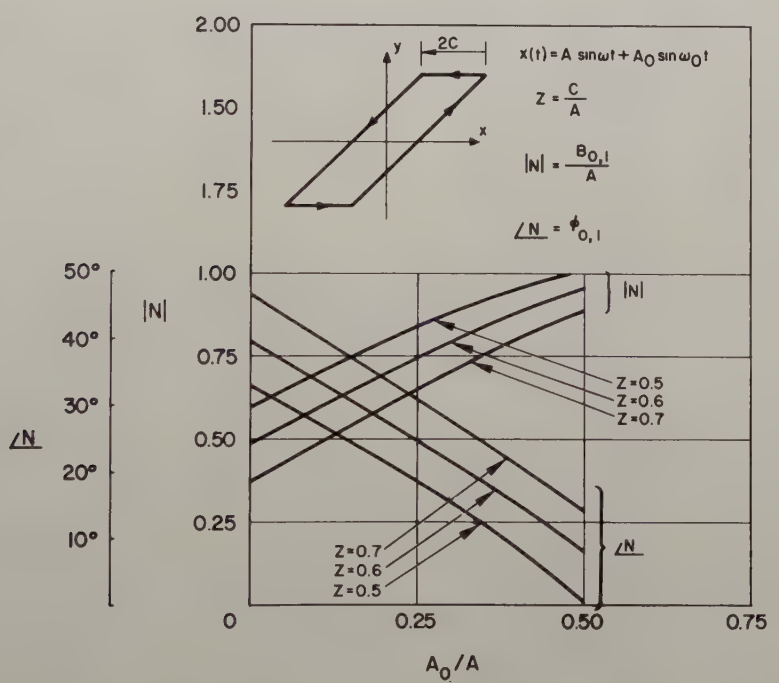


Fig. 11. The variation of the modified describing function (absolute value and phase) with respect to dither amplitude and signal level for a back-lash-type nonlinearity.

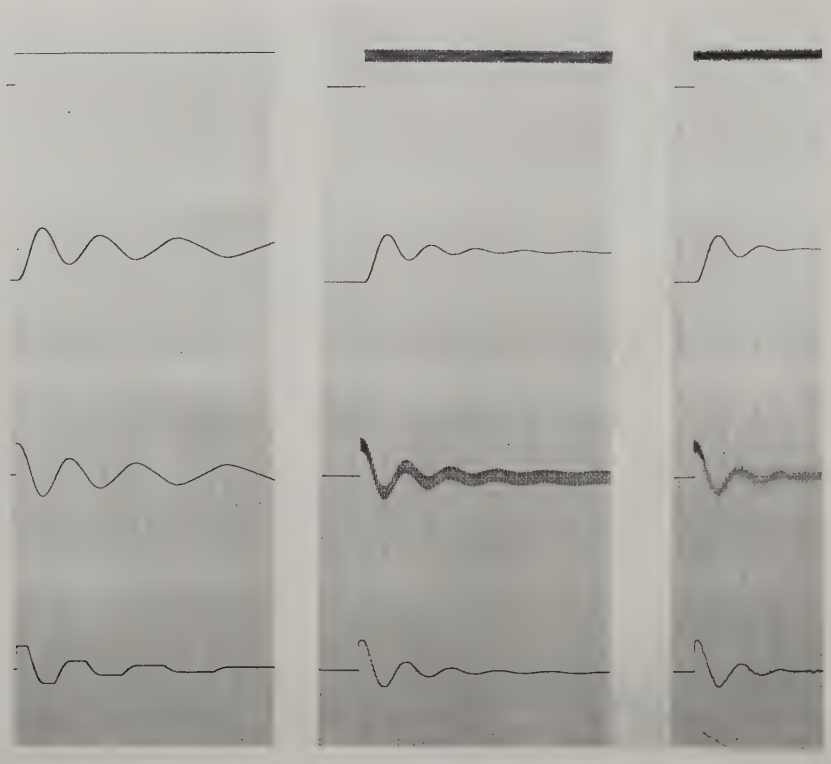


Fig. 12. Response of system containing backlash, (a) without dither, (b) with harmonic dither signal, (c) with noise dither.

```
      DIMENSION C(3), A(6), D(6, 3), FY(6, 3)
      DO8K=1, 3
      DD=K
      C(K)=0.4+DD/10.
      DO8J=1, 6
      DDD=J
      A(J)=DDD/10.-0.1
      SCM=0.
      SSM=0.
      L=0
      DO9I=1, 1818
      E=I
      X=COS1F(E*1.5707963/909.)+A(J)*COS1F(E*1.5707963/9.)
      IF(X-(1.+A(J)-2*C(K)))1, 2, 2
2     XX=1.+A(J)-C(K)
      L=L+1
      IF(I-L)3, 3, 4
1     IF((TT-C(K))-X) 4, 4, 5
5     XX=X        +C(K)
      GOTO3
4     XX=TT
3     YC=(1./909.)*XX*COS1F(E*1.5707963/909.)
      YS=(1./909.)*XX*SIN1F(E*1.5707963/909.)
      SCM=SCM+YC
      SSM=SSM+YS
9     TT=XX
      D(J, K)=SQRT1F(SCM**2.+SSM**2.)
      FY(J, K)=(90./1.5707963)*RTAN1F(SSM/SCM)
8     PRINT10, C(K), A(J), D(J, K), FY(J, K)
10    FORMAT (4F16.8)
      STOP777
```

Fig. 13. FORTRAN statement used in computation of curves in Fig. 11.

ON THE OPTIMUM SYNTHESIS OF SAMPLED DATA MULTIPOLE FILTERS WITH RANDOM AND NONRANDOM INPUTS*

H. C. Hsieh, C. T. Leondes
Department of Engineering,
University of California
Los Angeles 24, California

Summary

This paper considers the synthesis of optimum sampled data multipole filters with n inputs and m outputs. The signal portion of each input is assumed to consist of a stationary random component and a polynomial with unknown coefficients but known maximum order. Each signal is corrupted by stationary random noise. The filter under investigation is linear, time-invariant, and has finite memory. Each input to the filter consists of a sequence of impulses with a constant period T. Each impulse is assumed to have an area equal to the value of the signal plus noise at the sampling instant.

The synthesis procedure to be developed is to specify the weighting functions of the filter such that the system error, which is defined as the difference between the actual and ideal outputs, has zero ensemble mean and the system ensemble mean square error is minimum. The weighting functions thus obtained will have, in general, abrupt jumps at the sampling instants but they are continuous within the sampling intervals.

The synthesis procedure is extended to the case shown in Appendix A where each of the nonrandom signals can be expressed as an arbitrary linear combination of a set of known time functions. Further generalization is possible to the synthesis of time-varying filter with sampled nonstationary random inputs as given in Appendix B.

Introduction

A Survey of Two-pole Linear Filtering Theory

A two-pole system is defined as a device which has only one input terminal and one output terminal. During the past few years, considerable attention is given to the synthesis of filters whose signal contains both random and nonrandom components. Zadeh and Ragazzini[1] have solved the problem of synthesizing optimum, continuous, linear, time invariant filters, having finite memory time, for use with signals consisting of a stationary random process and a nonrandom polynomial with unknown coefficients but known maximum degree n. They assume the signal to be corrupted by an additive stationary random noise. The Zadeh-Ragazzini filter can be reduced to a Wiener[2] filter if the nonrandom part of the input is taken to be zero and the memory time is infinite. It has been shown that the optimum weighting function will involve impulse functions of various orders.

*This research was supported by the United States Air Force under Contract No. AF 49(638)-438 monitored by the Air Force Office Of Scientific Research of the Air Research and Development Command.

With the development of sampled data systems, statistical design techniques have been adapted to the case where the input is observed only at discrete instants of time. Franklin[3] considered the optimum synthesis of a sampled data filter and control system with input being the sum of a stationary random message and a stationary random noise. His problem is exactly the sampled data counterpart of Wiener's theory. For the optimum filter the transfer function is obtained by using the usual spectral shaping method. It will contain, in most cases, the product of two terms, one of which is a rational function of S and the other is a rational function of e^{ST}. The control system problem has also been investigated by Chang[4]. However, his synthesis procedure is carried out mostly in the Z-domain as contrasted to Franklin's time-domain approach.

The extension of the optimum filtering theory of Zadeh and Ragazzini to the sampled data case has been given by Lees[5]. His optimum linear filter will convert its input sequence of impulse functions into a smoothed output subject to the following conditions: the weighting function of the filter is not zero over a finite range only; in absence of random components of the input, the interpolation or extrapolation is error-free; in the presence of any random signal, the mean-square error of the output is minimized. It has been shown that the weighting function of this optimum filter will be piecewise continuous within the sampling intervals. Later on, Bergen[6] also considered the optimum linear filtering of a polynomial message plus stationary random noise. His investigation is very similar to that of Lees but his approach is more conventional.

Recent Development of Multipole Systems Theory

As stated in the first section, the two-pole filter problem has been intensively studied by previous investigators. Not until recently has the synthesis of multipole filters or control systems received due consideration. The authors[7,8] have developed the solution of optimum weighting functions in the Wiener sense for the multipole filter or control system where the input to each terminal consists of stationary random signal and noise. A set of integral equations is obtained by minimizing the mean-square errors of the system outputs. Through transformation, these integral equations can be converted into algebraic equations. For filter or free configuration control system, the optimum system transfer function can be found by solving these algebraic equations and using the method of undetermined coefficients. In the semi-free configuration control system, the

optimum transfer functions of compensation can be obtained in a similar manner such that the outputs of the fixed plant are more desirable.

With the increasing occurrence in practical problems of inputs available only at discrete instants of time, it is highly necessary to develop a theory for the optimum synthesis of sampled data multipole filter or free configuration control system. This paper presents the theoretical design equations for this important problem.

The Linear Filtering For Sampled Data Multipole Systems

Statement of the Filtering Problem

The filter under investigation is a sampled data multipole system with n inputs and m outputs. In order to facilitate the formulation of the problem, sampling switches are fictitiously inserted at the input terminals. The sampling process is then obtained by closing these switches briefly every T seconds such that the inputs are observed at these discrete instants of time.

The input $I_k(t)$ to each terminal before sampling is continuous time function and will consist of signal and noise with the following properties:

(1) The signal $S_k(t)$ has two components. One of them is a stationary random process $M_k(t)$ and the other one is a nonrandom function $P_k(t)$ which can be represented by a polynomial of known maximum degree m_k but with arbitrarily unknown coefficients.

(2) The noise $N_k(t)$ is also a stationary random process.

(3) It is assumed that the statistical characteristics of the random signals and noises are completely known.

With the above statements, the total signal to each terminal of the filter can be expressed by

$$S_k(t) = M_k(t) + P_k(t) \tag{1}$$

where

$$P_k(t) = \sum_{r=0}^{m_k} A_{kr}\, t^r \tag{2}$$

for k = 1, 2, ... n

The input $I_k(t)$ is then

$$I_k(t) = M_k(t) + P_k(t) + N_k(t) \tag{3}$$

It should be noted that the orders of the polynomial signals $P_k(t)$ can be different from terminal to terminal in order to give the best representation for the nonrandom inputs.

Because of the extremely short closing duration for the samples in comparison with the sampling interval T under actual physical situation and also because of the resultant mathematical simplifications, the sampling operation is taken as an impulse modulation of the continuous function. Thus the sampled input is

$$I_k^*(t) = I_k(t) \sum_{i=-\infty}^{+\infty} \delta(t-iT) \tag{4}$$

where $\delta(t)$ is an impulse function. The system block diagram is shown in Fig. 1. Here the W_{jk}'s are the weighting functions between the various input and output terminals.

The synthesis procedure to be developed is to specify all the system weighting functions satisfying the following conditions:

(1) Each system error, which is defined as the difference between the actual and ideal outputs, has zero ensemble mean; i.e.

$$\overline{\epsilon_j(t)} = 0 \tag{5}$$

for j = 1, 2, ... m

(2) Each ensemble mean square error of the system is a minimum with respect to the weighting functions W_{jk}'s which satisfy the condition 1.

$$\overline{\epsilon_j^2(t)} = \text{min. w.r.t. } W_{jk}\text{'s} \tag{6}$$

for j = 1, 2, ... m

(3) This multipole filter is physically realizable, linear, time invariant, and with finite memory. The last statement implies that the weighting functions W_{jk}'s should vanish for all time instants greater than a certain specified number. For simplicity, this number is chosen to be a multiple of the sampling interval T. Thus we have

$$W_{jk}(t) = 0 \qquad \text{for } t < 0 \text{ and } t > (N+1)T \tag{7}$$

where N is an integer.

It is evident that the synthesis approach is first to obtain the constraint equations imposed by condition (1) and (3), and then to find the optimum weighting functions which will minimize the system error subjected to these constraint equations. The filter is consequently considered to be the best under these design criteria.

Generation of System Error

The filtering problem to be discussed in this report is in a very broad sense where the ideal outputs can be any linear operation on the signal portions of the inputs. Let the ideal weighting functions for the system be $D_{jk}(t)$. Then the desired outputs are

$$C_{dj}(t) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau)\, S_k(t-\tau)\, d\tau \tag{8}$$

for j = 1, 2, ... m

The lower integration limit of "$-\infty$" emphasizes the fact that ideal weighting functions can be physically unrealizable.

The actual filter outputs can be expressed as

$$C_j(t) = \sum_{k=1}^{n} \int_{o}^{t_o} W_{jk}(\tau)\, I_k^*(t-\tau)\, d\tau \qquad (9)^*$$

where $W_{jk}(\tau) = 0$ for $\tau < 0$ and $\tau > t_o$. By definition, the system errors are

$$\epsilon_j(t) = C_j(t) - C_{dj}(t)$$

$$= \sum_{k=1}^{n} \int_{o}^{t_o} W_{jk}(\tau)\, I_k^*(t-\tau)\, d\tau$$

$$- \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau)\, S_k(t-\tau)\, d\tau \qquad (10)$$

The formulation of this problem will be best understood by referring to Fig. 2. For simplicity, only a single line flow diagram is drawn. It is apparent that the system outputs are compared with ideal outputs at all time so that the filter will be a continuous one.

Derivation of Constraint Equations

The constraint equations shall be obtained by applying the condition that the mean system error is zero. Combining Eqs. (1), (3), (4) and (10) gives the system error as

$$\epsilon_j(t) = \sum_{k=1}^{n} \int_{o}^{t_o} W_{jk}(\tau) I_k(t-\tau) \sum_{i=-\infty}^{+\infty} \delta(t-\tau-iT)\, d\tau$$

$$- \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau) S_k(t-\tau)\, d\tau$$

$$= \sum_{k=1}^{n} \int_{o}^{t_o} W_{jk}(\tau) \Big[M_k(t-\tau) + N_k(t-\tau) \Big]$$

$$\times \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)\, d\tau$$

$$- \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau) M_k(t-\tau)\, d\tau$$

$$+ \sum_{k=1}^{n} \Big[\int_{o}^{t_o} W_{jk}(\tau) P_k(t-\tau) \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)\, d\tau$$

$$- \int_{-\infty}^{\infty} D_{jk}(\tau) P_k(t-\tau)\, d\tau \Big] \qquad (11)$$

If it is assumed that the ensemble averages of the random signals and noises are zero

$$\overline{M_k(t)} = \overline{N_k(t)} = 0 \qquad (12)$$

it follows that

$$\overline{\epsilon_j(t)} = \sum_{k=1}^{n} \Big[\int_{o}^{t_o} W_{jk}(\tau) P_k(t-\tau) \sum_{i=-\infty}^{\infty} (t-\tau-iT)\, d\tau$$

$$- \int_{-\infty}^{\infty} D_{jk}(\tau) P_k(t-\tau)\, d\tau \Big] \qquad (13)$$

With $\overline{\epsilon_j(t)} = 0$, the following equation is generally obtained.

$$\int_{o}^{t_o} W_{jk}(\tau) P_k(t-\tau) \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)\, d\tau$$

$$= \int_{-\infty}^{\infty} D_{jk}(\tau) P_k(t-\tau)\, d\tau \qquad (14)$$

for $k = 1, 2, \ldots n$

It is evident that, with zero means for the random components of the inputs, the first condition imposed on the weighting functions can be rephrased as: In the absence of random components of the inputs, the actual filter outputs are identical to the ideal filter outputs.

The function $P_k(t-\tau)$ is now expanded into Taylor's series about $P_k(t)$. Since $P_k(t)$ is a polynomial of m_k degree, this expansion will give

$$P_k(t-\tau) = \sum_{r=o}^{m_k} \frac{(-1)^r}{r!} P_k^{(r)}(t)\, \tau^r \qquad (15)$$

By substituting Eq. (15) into Eq. (14), there

*The case where t_o differs for each input when a particular output is under consideration, is a relatively straightforward extension of the work presented here.

results

$$\sum_{r=o}^{m_k} \frac{(-1)^r}{r!} P_k^{(r)}(t) \int_o^{t_o} \tau^r W_{jk}(\tau) \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)d\tau$$

$$= \sum_{r=o}^{m_k} \frac{(-1)^r}{r!} P_k^{(r)}(t) \int_{-\infty}^{\infty} \tau^r D_{jk}(\tau)d\tau \qquad (16)$$

If this equality is true for all values of r, the following condition must exist.

$$\int_o^{t_o} \tau^r W_{jk}(\tau) \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)d\tau$$

$$= \int_{-\infty}^{\infty} \tau^r D_{jk}(\tau)d\tau \qquad (17)$$

$$\text{for } r = 0, 1, \ldots m_k$$

$$k = 1, 2, \ldots n$$

The left side of Eq. (17) needs further investigation. Since $W_{jk}(\tau)$ is also zero for $\tau > t_o$, the range for the index i is restricted. It can easily be verified that

$$\frac{t}{T} \geqq i \geqq \frac{t - t_o}{T} \qquad (18)$$

Define

$$u_{kr} = \int_o^{t_o} \tau^r W_{jk}(\tau) \sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)d\tau \qquad (19)$$

Then, by equation (18), it can be expressed in summation form as

$$u_{kr} = \sum_{i=\frac{t-t_o}{T}}^{\frac{t}{T}} (t-iT)^r W_{jk}(t-iT) \qquad (20)$$

Let t be replaced by

$$t = h\,T + \Delta \qquad (21)$$

where h is an integer and $0 \leqq \Delta < T$. Then Eq. (20) can be rewritten, with the further substitution of $t_o = NT + \Delta$, as

$$u_{kr} = \sum_{i=h-N}^{h+\frac{\Delta}{T}} (hT-iT+\Delta)^r W_{jk}(hT-iT+\Delta) \qquad (22)$$

If now h-i is replaced by q which must be an integer, it can easily be shown that

$$u_{kr} = \sum_{q=o}^{N} (qT+\Delta)^r W_{jk}(qT+\Delta) \qquad (23)$$

The range for the index q is quite consistent with the condition that $W_{jk}(t) = 0$ for $t < 0$ and $t > (N+1)T$ as it should be. It is evident that the weighting function $W_{jk}(t)$ is defined piecewise during each sampling interval.

The right side of Eq. (17) may be denoted by

$$v_{kr} = \int_{-\infty}^{\infty} \tau^r D_{jk}(\tau)d\tau \qquad (24)$$

where v_{kr} is a constant.

Then Eq. (17) can be stated more conveniently by

$$u_{kr} = v_{kr} \qquad (25)$$

$$\text{for } k = 1, 2, \ldots n$$

$$r = 0, 1, \ldots m_k$$

These are $\sum_{k=1}^{n} m_k + n$ constraint equations which $W_{jk}(qT+\Delta)$ must satisfy for all values of Δ.

By inspection of Eqs. (23) and (24) it may be noted that, in order to obtain error-free outputs in the absence of random components of the inputs and to minimize the mean square errors in the presence of them, the number N for the filter must be greater than the maximum order of all the polynomial inputs. In other words, the filter must be able to remember more than $m_k + 1$ previous sample values.

Optimum Synthesis of Sampled Data Multipole Filter

Minimization of the Mean Square Error

Suppose now that the constraints given by Eq. (25) are satisfied, the expression for error will be reduced to

$$\epsilon_j(t) = \sum_{k=1}^{n} \int_o^{t_o} W_{jk}(\tau)\left[M_k(t-\tau)+N_k(t-\tau)\right] \times \sum_{j=-\infty}^{\infty} \delta(t-\tau-jT)d\tau$$

$$- \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau)M_k(t-\tau)d\tau \qquad (26)$$

Let t be replaced by $hT + \Delta$ as before. Then, by using the same manipulation as in Eqs. (22) and

(23), it is easily shown that

$$\epsilon_j(t) = \epsilon_j(hT + \Delta)$$

$$= \sum_{k=1}^{n} \int_0^{\infty} W_{jk}(\tau) \left[M_k(t-\tau) + N_k(t-\tau) \right] \times \sum_{q=0}^{N} \delta(\tau - qT - \Delta) d\tau$$

$$- \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau) M_k(t-\tau) d\tau \qquad (27)$$

The mean square error is then given by

$$\overline{\epsilon_j^2(t)} = \sum_{k=1}^{n} \sum_{k'=1}^{n} \int_0^{\infty} W_{jk'}(\tau_2) \sum_{\ell=0}^{N} \delta(\tau_2 - \ell T - \Delta) d\tau_2 \times \int_0^{\infty} W_{jk}(\tau_1) \phi_{k'k}(\tau_2 - \tau_1) \sum_{q=0}^{N} \delta(\tau_1 - qT - \Delta) d\tau_1$$

$$-2 \sum_{k=1}^{n} \sum_{k'=1}^{n} \int_0^{\infty} W_{jk'}(\tau_2) \sum_{\ell=0}^{N} \delta(\tau_2 - \ell T - \Delta) d\tau_2 \times \int_{-\infty}^{\infty} D_{jk}(\tau_1) \theta_{k'k}(\tau_2 - \tau_1) d\tau_1$$

$$+ \sum_{k=1}^{n} \sum_{k'=1}^{n} \int_{-\infty}^{\infty} D_{jk'}(\tau_2) d\tau_2 \times \int_{-\infty}^{\infty} D_{jk}(\tau_1) \phi_{M_{k'}M_k}(\tau_2 - \tau_1) d\tau_1 \qquad (28)$$

where $\phi_{k'k}$ and $\theta_{k'k}$ are the summation of various autocorrelation functions and cross-correlation functions for random signals and noises as defined by

$$\phi_{k'k}(\tau) = \phi_{M_{k'}M_k}(\tau) + \phi_{M_{k'}N_k}(\tau) + \phi_{N_{k'}M_k}(\tau) + \phi_{N_{k'}N_k}(\tau) \qquad (29)$$

$$\theta_{k'k}(\tau) = \phi_{M_{k'}M_k}(\tau) + \phi_{N_{k'}M_k}(\tau) \qquad (30)$$

The above correlation functions are the ensemble averages of the random variables. If the ergodic property is applicable to these random processes, these ensemble averages can be equated to the averages with respect to time performed on a single representative function of the ensemble.

The next step in the development is to minimize the mean square error with respect to the weighting functions subjected to the constraints in Eq. (25). By introducing the Lagrangian multiplier λ_{kr}, this problem is equivalent to minimizing the expression

$$Z_j(W) = \overline{\epsilon_j^2(t)} + 2 \sum_{k'=1}^{n} \sum_{r=0}^{m_{k'}} \lambda_{k'r} u_{k'r}$$

$$= \sum_{k=1}^{n} \sum_{k'=1}^{n} \int_0^{\infty} W_{jk'}(\tau_2) \sum_{\ell=0}^{N} \delta(\tau_2 - \ell T - \Delta) d\tau_2 \times \int_0^{\infty} W_{jk}(\tau_1) \phi_{k'k}(\tau_2 - \tau_1) \sum_{q=0}^{N} \delta(\tau_1 - qT - \Delta) d\tau_1$$

$$-2 \sum_{k'=1}^{n} \int_0^{\infty} W_{jk'}(\tau_2) \sum_{\ell=0}^{N} \delta(\tau_2 - \ell T - \Delta) d\tau_2 \times \left[\sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau_1) \theta_{k'k}(\tau_2 - \tau_1) d\tau_1 - \sum_{r=0}^{m_{k'}} \lambda_{k'r} \tau_2^r \right]$$

$$+ \sum_{k=1}^{n} \sum_{k'=1}^{n} \int_{-\infty}^{\infty} D_{jk'}(\tau_2) d\tau_2 \times \int_{-\infty}^{\infty} D_{jk}(\tau_1) \phi_{M_{k'}M_k}(\tau_2 - \tau_1) d\tau_1 \qquad (31)$$

The technique[7,8] employed in minimizing Eq. (31) is essentially to complete the square for the first two summation terms. Let the arbitrary functions $U_{jk}(\tau)$'s be defined implicitly by the following equations:

$$\sum_{k=1}^{n} \int_0^{\infty} U_{jk}(\tau_1) \phi_{k'k}(\tau_2 - \tau_1) \sum_{q=0}^{N} \delta(\tau_1 - qT - \Delta) d\tau_1$$

$$= \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(\tau_1)\theta_{k'k}(\tau_2-\tau_1)d\tau_1 - \sum_{r=0}^{m_{k'}} \lambda_{k'r}\tau_2^r$$

$$\text{for } k' = 1, 2, \ldots n \qquad (32)$$

By substituting Eq. (32) into Eq. (31) and adding and substracting a term

$$\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{0}^{\infty} U_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\int_{0}^{\infty} U_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

there results

$$Z_j(W) = \sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{0}^{\infty} W_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\int_{0}^{\infty} W_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$-2\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{0}^{\infty} W_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\int_{0}^{\infty} U_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$+\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{0}^{\infty} U_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\int_{0}^{\infty} U_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$-\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{0}^{\infty} U_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\int_{0}^{\infty} U_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$+\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_1)d\tau_1$$

$$\times\int_{-\infty}^{\infty} D_{jk'}(\tau_2)\phi_{M_{k'}M_k}(\tau_2-\tau_1)d\tau_2 \qquad (33)$$

It is quite evident that the first three summation terms can be expressed as

$$\left\{\sum_{k=1}^{n}\int_{0}^{\infty}\Big[W_{jk}(\tau)-U_{jk}(\tau)\Big]\Big[M_k(t-\tau)+N_k(t-\tau)\Big]\right.$$

$$\left.\times\sum_{q=0}^{N}\delta(\tau-qT-\Delta)d\tau\right\}^2 \qquad (34)$$

which shows the fact that this term is always positive. Thus, to minimize Eq. (33), it is obvious that the following choice can be made

$$U_{jk}(\tau) = W_{jk}(\tau) \qquad (35)$$

Under this condition, Eq. (32) becomes

$$\sum_{k=1}^{n}\int_{0}^{\infty} W_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$= \sum_{k=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_1)\theta_{k'k}(\tau_2-\tau_1)d\tau_1 - \sum_{r=0}^{m_{k'}} \lambda_{k'r}\tau_2^r$$

$$\text{for } k' = 1, 2, \ldots n$$

$$\tau_2 = \ell T + \Delta \text{ where } = 0, 1, \ldots N \qquad (36)$$

To show that Eq. (36) gives not only the necessary condition for optimum system but also the sufficient one, let the optimum weighting functions $W_{jk}(\tau)$ be replaced by another arbitrary weighting function of the form $W_{jk}(\tau) - Y_{jk}(\tau)$. Then Eq. (31) becomes

$$Z_j' = Z_j + 2\sum_{k'=1}^{n}\int_0^{\infty} Y_{jk'}(\tau_2)\sum_{\ell=0}^{N}\delta(\tau_2-\ell T-\Delta)d\tau_2$$

$$\times\left[\sum_{k=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_1)\theta_{k'k}(\tau_2-\tau_1)d\tau_1-\sum_{r=0}^{m_{k'}}\lambda_{k'r}\tau_2^{r}\right.$$

$$\left.-\sum_{k=1}^{n}\int_0^{\infty} W_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1\right]$$

$$+\left\{\sum_{k=1}^{n}\int_0^{\infty} Y_{jk}(\tau)\left[M_k(t-\tau)+N_k(t-\tau)\right]\right.$$

$$\left.\times\sum_{q=0}^{N}\delta(\tau-qT-\Delta)d\tau\right\}^2 \qquad (37)$$

Thus, if Eq. (36) is satisfied, Z_j' reduces to

$$Z_j' = Z_j +\left\{\sum_{k=1}^{n}\int_0^{\infty} Y_{jk}(\tau)\left[M_k(t-\tau)+N_k(t-\tau)\right]\right.$$

$$\left.\times\sum_{q=0}^{N}\delta(\tau-qT-\Delta)d\tau\right\}^2 \qquad (38)$$

Hence, it is seen that the value of Z_j is not decreased by any perturbation of W_{jk} if Eq. (36) is fulfilled. The sufficient condition is thus established.

Solution for the Optimum Weighting Functions

In order to solve the optimum weighting functions, Eq. (36) is rewritten for investigation.

$$\sum_{k=1}^{n}\int_0^{\infty} W_{jk}(\tau_1)\phi_{k'k}(\tau_2-\tau_1)\sum_{q=0}^{N}\delta(\tau_1-qT-\Delta)d\tau_1$$

$$=\sum_{k=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_1)\theta_{k'k}(\tau_2-\tau_1)d\tau_1$$

$$-\sum_{r=0}^{m_{k'}}\lambda_{k'r}\tau_2^{r} \qquad (39)$$

for $k' = 1, 2, \ldots n$

$\tau_2 = \ell T + \Delta$ where $\ell = 0, 1, \ldots N$

Define

$$\sum_{k=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_1)\theta_{k'k}(\tau_2-\tau_1)d\tau_1 = B_{jk'}(\tau_2) \qquad (40)$$

Eq. (39) can be rewritten in summation form as

$$\sum_{k=1}^{n}\sum_{q=0}^{N}\phi_{k'k}(\ell T-qT)\,W_{jk}(qT+\Delta)$$

$$= B_{jk'}(\ell T+\Delta)-\sum_{r=0}^{m_{k'}}\lambda_{k'r}(\ell T+\Delta)^{r} \qquad (41)$$

for $k' = 1, 2, \ldots n$

$\ell = 0, 1, \ldots N$

$0 \leqq \Delta < T$

These are n x (N + 1) simultaneous equations which can be used to solve for $W_{jk}(qT+\Delta)$ in terms of $\lambda_{k'r}$. By using the $\sum_{k=1}^{n} m_k + n$ constraints given by Eq. (25), these Lagrangian multipliers $\lambda_{k'r}$ will be determined. Thus, combination of Eqs (25) and (41) gives a unique solution for the optimum weighting functions of the filter.

For convenience, let Eqs. (41) and (25) be rewritten as

$$\sum_{k=1}^{n}\sum_{q=0}^{N}\phi_{k'k}(\ell T-qT)W_{jk}(qT+\Delta)+\sum_{r=0}^{m_{k'}}\lambda_{k'r}(\ell T+\Delta)^{r}$$

$$= B_{jk'}(\ell T+\Delta) \qquad (42A)$$

for $k' = 1, 2, \ldots n$

$\ell = 0, 1, \ldots N$

and

$$\sum_{q=0}^{N}(qT+\Delta)^{r}\,W_{jk}(qT+\Delta) = v_{kr} \qquad (42B)$$

for $k = 1, 2, \ldots n$

$r = 0, 1, \ldots m_k$

In order to solve Eq. (42) by matrix algebra, it will be desirable to assume that the polynomial inputs have the same order m_p. For any polynomial input which has order less than m_p, its corresponding Lagrangian multipliers $\lambda_{k'r}$ with subscript $r > m_{k'}$ will be zero. In addition, Eq. (42B) will

be identically zero for these subscripts. Thus the total equations to be considered now are

$$n \times \left[(N+1) + (m_p + 1) \right].$$

Let the following submatrices be defined:

$$\underline{E_{jk'}} = \begin{bmatrix} B_{jk'}(\Delta) \\ B_{jk'}(T+\Delta) \\ \vdots \\ B_{jk'}(NT+\Delta) \end{bmatrix} = (N+1) \text{ column vector} \qquad (43)$$

$$\underline{V_{jk}} = \begin{bmatrix} v_{k0} \\ v_{k1} \\ \vdots \\ v_{km_p} \end{bmatrix} = (m_p+1) \text{ column vector} \qquad (44)$$

$$\underline{F_{jk}} = \begin{bmatrix} W_{jk}(\Delta) \\ W_{jk}(T+\Delta) \\ \vdots \\ W_{jk}(NT+\Delta) \end{bmatrix} = (N+1) \text{ column vector} \qquad (45)$$

$$\underline{G_{jk'}} = \begin{bmatrix} \lambda_{k'0} \\ \lambda_{k'1} \\ \vdots \\ \lambda_{k'm_p} \end{bmatrix} = (m_p+1) \text{ column vector} \qquad (46)$$

$$\underline{H_{k'k}} = \begin{bmatrix} \phi_{k'k}(0,0)\phi_{k'k}(0,1)\ldots\phi_{k'k}(0,N) \\ \phi_{k'k}(1,0)\phi_{k'k}(1,1)\ldots\phi_{k'k}(1,N) \\ \vdots \\ \phi_{k'k}(N,0)\phi_{k'k}(N,1)\ldots\phi_{k'k}(N,N) \end{bmatrix} = (N+1) \times (N+1) \text{ square matrix (Symmetric matrix)} \qquad (47)$$

where $\phi_{k'k}(\ell,q) \triangleq \phi_{k'k}(\ell T - qT)$.

$$\underline{I_{k'}} = \begin{bmatrix} \Delta^0 & \Delta^1 & \ldots & \Delta^{m_p} \\ (T+\Delta)^0 & (T+\Delta)^1 & \ldots & (T+\Delta)^{m_p} \\ \vdots & \vdots & & \vdots \\ (NT+\Delta)^0 & (NT+\Delta)^1 & \ldots & (NT+\Delta)^{m_p} \end{bmatrix} = (N+1) \times (m_p+1) \text{ matrix} \qquad (48)$$

and

$$\underline{J_k} = \underline{I'_{k'}} = \begin{bmatrix} \Delta^0 & (T+\Delta)^0 & \ldots & (NT+\Delta)^0 \\ \Delta^1 & (T+\Delta)^1 & \ldots & (NT+\Delta)^1 \\ \vdots & \vdots & & \vdots \\ \Delta^{n'_p} & (T+\Delta)^{m_p} & \ldots & (NT+\Delta)^{m_p} \end{bmatrix} = (m_p+1) \times (N+1) \text{ matrix} \qquad (49)$$

It should be noted that, with the actual order of any polynomial input less than the maximum order m_p, the appropriate elements in submatrices $\underline{G_{jk'}}$, $\underline{I_{k'}}$, $\underline{J_k}$, and $\underline{V_{jk}}$ will be zero.

Then Eq. (42) can be rewritten in normalized matrix form as

$$\left[\begin{array}{cccc|cccc} \underline{H}_{11} & \underline{H}_{12} & \cdots & \underline{H}_{1n} & \underline{I}_1 & \underline{0} & \cdots & \underline{0} \\ \underline{H}_{21} & \underline{H}_{22} & \cdots & \underline{H}_{2n} & \underline{0} & \underline{I}_2 & \cdots & \underline{0} \\ \vdots & \vdots & & \vdots & \vdots & & & \vdots \\ \underline{H}_{n1} & & & \underline{H}_{nn} & \underline{0} & \underline{0} & \cdots & \underline{I}_n \\ \hline \underline{J}_1 & \underline{0} & \cdots & \underline{0} & & & & \\ \underline{0} & \underline{J}_2 & \cdots & \underline{0} & & & & \\ \vdots & \vdots & & \vdots & & & & \\ \underline{0} & \underline{0} & \cdots & \underline{J}_n & & & & \end{array}\right] \left[\begin{array}{c} \underline{F}_{j1} \\ \underline{F}_{j2} \\ \vdots \\ \underline{F}_{jn} \\ \hline \underline{G}_{j1} \\ \underline{G}_{j2} \\ \vdots \\ \underline{G}_{jn} \end{array}\right] = \left[\begin{array}{c} \underline{E}_{j1} \\ \underline{E}_{j2} \\ \vdots \\ \underline{E}_{jn} \\ \hline \underline{V}_{j1} \\ \underline{V}_{j2} \\ \vdots \\ \underline{V}_{jn} \end{array}\right] \qquad (50)$$

Eq. (50) will be expressed in a more compact form by further defining

$$\underline{H} = \begin{bmatrix} \underline{H}_{11} & \underline{H}_{12} & \cdots & \underline{H}_{1n} \\ \underline{H}_{21} & \underline{H}_{22} & \cdots & \underline{H}_{2n} \\ \vdots & \vdots & & \vdots \\ \underline{H}_{n1} & \underline{H}_{n2} & \cdots & \underline{H}_{nn} \end{bmatrix} = \text{square matrix} \qquad (51)$$

$$\underline{I} = \begin{bmatrix} \underline{I}_1 & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{I}_2 & \cdots & \underline{0} \\ \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{I}_n \end{bmatrix} \qquad (52)$$

$$\underline{J} = \begin{bmatrix} \underline{J}_1 & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{J}_2 & \cdots & \underline{0} \\ \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{J}_n \end{bmatrix} \qquad (53)$$

$$\underline{0} = \begin{bmatrix} 0 \end{bmatrix} = \text{null matrix} \qquad (54)$$

$$\underline{F}_j = \begin{bmatrix} \underline{F}_{j1} \\ \underline{F}_{j2} \\ \vdots \\ \underline{F}_{jn} \end{bmatrix} = \text{column vector} \qquad (55)$$

$$\underline{G}_j = \begin{bmatrix} \underline{G}_{j1} \\ \underline{G}_{j2} \\ \vdots \\ \underline{G}_{jn} \end{bmatrix} = \text{column vector} \qquad (56)$$

$$\underline{E}_j = \begin{bmatrix} \underline{E}_{j1} \\ \underline{E}_{j2} \\ \vdots \\ \underline{E}_{jn} \end{bmatrix} = \text{column vector} \qquad (57)$$

and

$$\underline{V}_j = \begin{bmatrix} \underline{V}_{j1} \\ \underline{V}_{j2} \\ \vdots \\ \underline{V}_{jn} \end{bmatrix} = \text{column vector} \qquad (58)$$

Thus it is obtained that

$$\begin{bmatrix} \underline{H} & \underline{I} \\ \underline{J} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{F}_j \\ \underline{G}_j \end{bmatrix} = \begin{bmatrix} \underline{E}_j \\ \underline{V}_j \end{bmatrix} \qquad (59)$$

Eq. (59) may be solved to yield

$$\begin{bmatrix} \underline{F}_j \\ \underline{G}_j \end{bmatrix} = \begin{bmatrix} \underline{R} & \underline{S} \\ \underline{X} & \underline{Y} \end{bmatrix} \begin{bmatrix} \underline{E}_j \\ \underline{V}_j \end{bmatrix} \tag{60}$$

where

$$\underline{R} = \underline{H}^{-1} - \underline{H}^{-1}\,\underline{I}\,(\underline{J}\,\underline{H}^{-1}\,\underline{I})\quad \underline{J}\,\underline{H}^{-1} \tag{61}$$

$$\underline{S} = \underline{H}^{-1}\,\underline{I}\,(\underline{J}\,\underline{H}^{-1}\,\underline{I})^{-1} \tag{62}$$

$$\underline{X} = (\underline{J}\,\underline{H}^{-1}\,\underline{I})^{-1}\,\underline{J}\,\underline{H}^{-1} \tag{63}$$

$$\underline{Y} = -(\underline{J}\,\underline{H}^{-1}\,\underline{I})^{-1} \tag{64}$$

Hence the weighting functions vector is given by

$$\underline{F}_j = \underline{R}\,\underline{E}_j + \underline{S}\,\underline{V}_j \tag{65}$$

$$\text{for } j = 1, 2, \ldots m$$

It is evident that the submatrices $\underline{R}$ and $\underline{S}$ are the same for all output terminals. However, the $\underline{E}_j$ and $\underline{V}_j$ submatrices will be a function of the particular output terminal concerned.

Although Eq. (65) gives a concise expression for the optimum weighting functions matrix, yet the evaluation of $\underline{R}$ and $\underline{S}$ as given by Eqs. (61) and (62) may actually be very tedious. It can be seen that these functions have, in general, abrupt jumps at the sampling instants but they are continuous within the sampling intervals. If Δ is equal to zero, a discrete filter is obtained.

Mean Square Error of Optimum Filter

The mean square error of a particular output terminal for the filter is given by Eq. (28). It can be rewritten as

$$\begin{aligned}
\overline{\epsilon_j^2(t)} = \overline{\epsilon_j^2(\Delta)} &= \sum_{k=1}^{n}\sum_{k'=1}^{n}\sum_{\ell=0}^{N}\sum_{q=0}^{N} W_{jk'}(\ell T+\Delta)\, W_{jk}(qT+\Delta)\,\phi_{k'k}(\ell T-qT) \\
&\quad -2\sum_{k=1}^{n}\sum_{k'=1}^{n}\sum_{\ell=0}^{N} W_{jk'}(\ell T+\Delta)\int_{-\infty}^{\infty} D_{jk}(\tau_1)\,\theta_{k'k}(\ell T+\Delta-\tau_1)\,d\tau_1 \\
&\quad +\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{-\infty}^{\infty} D_{jk}(\tau_2)\,d\tau_2 \int_{-\infty}^{\infty} D_{jk'}(\tau_1)\,\phi_{M_{k'},M_k}(\tau_2-\tau_1)\,d\tau_1 \\
&= \sum_{k=1}^{n}\sum_{k'=1}^{n}\sum_{\ell=0}^{N}\sum_{q=0}^{N} W_{jk'}(\ell T+\Delta)\, W_{jk}(qT+\Delta)\,\phi_{k'k}(\ell T-qT) \\
&\quad -2\sum_{k'=1}^{n}\sum_{\ell=0}^{N} W_{jk'}(\ell T+\Delta)\, B_{jk'}(\ell T+\Delta) \\
&\quad +\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{-\infty}^{\infty} D_{jk'}(\tau_2)\,d\tau_2 \int_{-\infty}^{\infty} D_{jk}(\tau_1)\,\phi_{M_{k'},M_k}(\tau_2-\tau_1)\,d\tau_1
\end{aligned} \tag{66}$$

where $B_{jk'}(\tau)$ is defined by Eq. (40). Since the random signals and noises are stationary and the sampled inputs are non-stationary, the mean square error is a function of Δ. An average of the mean square error over $0 \leqq \Delta < T$ will yield the average over all time.

The minimum mean square error will be obtained if the solution for optimum weighting functions as given by Eq. (65) is substituted in Eq. (66). By using the matrix notation, it can be conveniently expressed as

$$\begin{aligned}
\overline{\epsilon_j^2(\Delta)}_{min} &= \underline{F}_j'\,\underline{H}\,\underline{F}_j - 2\,\underline{E}_j'\,\underline{F}_j + A_j \\
&= \left\{ (\underline{R}\,\underline{E}_j + \underline{S}\,\underline{V}_j)'\,\underline{H} - 2\,\underline{E}_j' \right\} \times (\underline{R}\,\underline{E}_j + \underline{S}\,\underline{V}_j) + A_j
\end{aligned} \tag{67}$$

$$\text{for } j = 1, 2, \ldots m$$

Here the prime denotes the transpose and A_j is defined by

$$A_j = \sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{-\infty}^{\infty} D_{jk'}(\tau_2)\,d\tau_2$$

$$\times \int_{-\infty}^{\infty} D_{jk}(\tau_1)\phi_{M_k,M_k}(\tau_2-\tau_1)d\tau_1 \qquad (68)$$

An Illustrative Example

As a simple example, let a filter with two input-terminals and one output-terminal be considered. The signals to these two channels are assumed to be the same. However, they are corrupted by different noises. The polynomial inpus are taken to be of first order. The random signals and noises are not correlated. Furthermore, the noises are uncorrelated through the sampling interval T. The function of the filter is to extract the signal from these two different channels. The system block diagram is shown in Fig. 3.

It is evident that, in this problem,

$$M_1(t) = M_2(t) = M(t) \qquad (69)$$

$$P_1(t) = P_2(t) = P(t) = A_o + A_1 t \qquad (70)$$

Let the autocorrelation function of the random signal M(t) be

$$\phi_{MM}(\tau) = e^{-|\tau|} \qquad (71)$$

Under the assumption of uncorrelated noise samples through the interval T, the two autocorrelation functions for noises are given by

$$\phi_{N_1N_1}(\ell T-qT) = 0.5\,\delta_{\ell q} \qquad (72)$$

$$\phi_{N_2N_2}(\ell T-qT) = 0.2\,\delta_{\ell q} \qquad (73)$$

where $\delta_{\ell q}$ is Kronecker delta which is either one or zero according as the subscripts ℓ and q are or are not the same. In addition,

$$\phi_{MN_1}(\tau) = \phi_{MN_2}(\tau) = 0 \qquad (74)$$

The sampling interval is taken to be one second and the memory of the filter, three seconds, (N = 2).

The ideal weighting functions of the filter are

$$D_{11}(\tau) = D_{12}(\tau) = \delta(\tau) \qquad (75)$$

The overall correlation functions defined by Eqs. (29) and (30) are

$$\phi_{11}(\tau) = \phi_{MM}(\tau) + \phi_{N_1N_1}(\tau)$$

$$\phi_{12}(\tau) = \phi_{21}(\tau) = \phi_{MM}(\tau)$$

$$\phi_{22}(\tau) = \phi_{MM}(\tau) + \phi_{N_2N_2}(\tau)$$

$$\theta_{11}(\tau) = \theta_{12}(\tau) = \theta_{21}(\tau) = \theta_{22}(\tau) = \phi_{MM}(\tau)$$

Let the following submatrices be formed:

$$\underline{E_{11}} = 2\,e^{-|\Delta|}\begin{bmatrix} 1 \\ e^{-1} \\ e^{-2} \end{bmatrix} = \underline{E_{12}}$$

$$\underline{V_{11}} = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} = \underline{V_{12}}$$

$$\underline{H_{11}} = \begin{bmatrix} 1.5 & e^{-1} & e^{-2} \\ e^{-1} & 1.5 & e^{-1} \\ e^{-2} & e^{-1} & 1.5 \end{bmatrix}$$

$$\underline{H_{12}} = \begin{bmatrix} 1 & e^{-1} & e^{-2} \\ e^{-1} & 1 & e^{-1} \\ e^{-2} & e^{-1} & 1 \end{bmatrix} = \underline{H_{21}}$$

$$\underline{H_{22}} = \begin{bmatrix} 1.2 & e^{-1} & e^{-2} \\ e^{-1} & 1.2 & e^{-1} \\ e^{-2} & e^{-1} & 1.2 \end{bmatrix}$$

$$\underline{I_1} = \begin{bmatrix} 1 & \Delta \\ 1 & 1+\Delta \\ 1 & 2+\Delta \end{bmatrix} = \underline{I_2}$$

$$\underline{J_1} = \begin{bmatrix} 1 & 1 & 1 \\ \Delta & 1+\Delta & 2+\Delta \end{bmatrix} = \underline{J_2}$$

Even in this simple example, it is difficult to calculate the submatrices R and S in terms of Δ as given by Eqs. (61) and (62).

$$\underline{R} = \underline{H}^{-1} - \underline{H}^{-1}\,\underline{I}\,(\underline{J}\,\underline{H}^{-1}\,\underline{I})^{-1}\,\underline{J}\,\underline{H}^{-1}$$

$$\underline{S} = \underline{H}^{-1}\,\underline{I}\,(\underline{J}\,\underline{H}^{-1}\,\underline{I})$$

However, by using digital computer, the numerical values for the weighting functions can be easily obtained for every assumed value of Δ.

Conclusion

A synthesis procedure has been developed for obtaining the optimum sampled data multipole system with random and nonrandom inputs. With the assumption that the nonrandom signal components can be represented by polynomials, the solution for these optimum weighting functions becomes a simple routine matter. Theoretically, an analytic expression can be obtained for each weighting function in terms of Δ over each sampling interval T. However, due to the inherent complexity of multipole system, the matrix operation for obtaining the solution becomes very tedious even in the simplest case if Δ is carried all the way through as an independent variable. Thus digital computer is advised to be used so that the numerical answers can easily be found for every assumed value of Δ .

References

1. Zadeh, L. A., and Ragazzini, J. R.: "An Extension of Wiener's Theory of Prediction," J. Of Appl. Phys., vol. 21, No. 7, pp. 645-655, July 1950.

2. Wiener, N.: "The Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley & Sons, Inc., New York, 1949.

3. Franklin, G. F.: "The Optimum Synthesis of Sampled Data Systems," Electronics Research Laboratories, Columbia University, Technical Report T-61B, May 1955.

4. Chang, S. S. L.: "Statistical Design Theory for Digital Controlled Continuous Systems," Trans. AIEE, vol. 77, Pt. II, pp. 191-198, September 1958.

5. Lees, A. B.: "Interpolation and Extrapolation of Sampled Data," Trans. IRE on Information Theory, vol. IT-2, pp. 12-17, March 1956.

6. Bergen, A. R.: "Optimum Filtering of Sampled Polynomial Message Plus Random Noise," Electronics Research Laboratories, Columbia University Technical Report T-1/133, May 1950.

7. Hsieh, H. C., and C. T. Leondes: "On the Optimum Synthesis of Multipole Control Systems in The Wiener Sense," IRE, National Convention-al Record, vol. 7, part 4, pp. 18-31, March 1959.

8. Hsieh, H. C., and C. T. Leondes: "Techniques for the Optimum Synthesis of Multipole Control Systems with Random Processes as Inputs," presented at the National Automatic Control Conference, November, 1959.

Appendix A

Filtering With Generalized Nonrandom Inputs

In the previous discussion, assumption is made that the nonrandom signals $P_k(t)$ can be expressed by polynomials with known maximum degrees m_k but unknown coefficients A_{kr}. It is seen that the function $P_k(t-\tau)$ expanded into Taylor's series about $P_k(t)$ as given by Eq. (15) is a polynomial in τ with maximum degree of m_k and coefficients as functions of t. Hence V_{kr} obtained from Eq. (24) are constants. The continuous filter thus obtained is time invariant.

Extension of the solution can be made to the case where the nonrandom signals are not restricted to polynomials of degree m_k but can be expressed as an arbitrary linear combination of a set of known time functions. Thus, the nonrandom signals are given by

$$P_k(t) = \sum_{r=o}^{m_k} A_{kr}\, p_{kr}(t) \qquad \text{(A-1)}$$

where $p_{kr}(t)$ are known but A_{kr} are unknown. For a considerably large class of functions $p_{kr}(t)$, they have the property that $p_{kr}(t-\tau)$ can be represented in the form

$$p_{kr}(t-\tau) = \sum_{\alpha=1}^{m_{kr}} a_{kr}^{\alpha}(\tau)\, b_{kr}^{\alpha}(t) \qquad \text{(A-2)}$$

Under this condition, the constraint equations are

$$\sum_{q=o}^{N} a_{kr}^{\alpha}(qT+\Delta) W_{jk}(qT+\Delta) = \int_{-\infty}^{\infty} a_{kr}^{\alpha}(\tau) D_{jk}(\tau)\, d\tau$$

$$\text{for } k = 1, 2, \ldots n$$

$$r = 0, 1, \ldots m_k \qquad \text{(A-3)}$$

$$\alpha = 1, 2, \ldots m_{kr}$$

The simultaneous equations obtained in minimizing the mean square error are

$$\sum_{k=1}^{n} \sum_{q=o}^{N} \phi_{k'k}(\ell T - qT) W_{jk}(qT+\Delta) + \sum_{r=o}^{m_{k'}} \sum_{\alpha=1}^{m_{k'r}} \lambda_{k'r}^{\alpha}\, a_{k'r}^{\alpha}(\ell T+\Delta)$$

$$= B_{jk'}(\ell T+\Delta)$$

$$\text{for } k' = 1, 2, \ldots n$$

$$\ell = 0, 1, \ldots N \qquad \text{(A-4)}$$

$$0 \leqq \Delta < T$$

It is evident that, except slight modification of the matrices $\underline{I}$ and $\underline{J}$, the techniques developed previously are equally applicable to this generalized nonrandom inputs problem.

Appendix B

Time-Varying Filters With Sampled Nonstationary Random Inputs

The optimization theory thus far developed is based upon the conditions that the random inputs are stationary and the ideal weight functions are time-invariant. During recent years, the analysis and synthesis of time-variant systems have received considerable attention such as those systems in the guided-missile problem. Thus the desired outputs of the multipole filter can be time-varying linear operation upon the signal components of the inputs. In many practical problems, the assumption that the random inputs are stationary is not valid. The nonstationary random inputs imply that their statistical properties varying with time. Therefore, the synthesis of time varying filters with sampled nonstationary random inputs need to be considered.

For this more general problem, the nonrandom inputs will be assumed of the form expressed by Eq. (A-1).

$$P_k(t) = \sum_{r=o}^{m_k} A_{kr}\, p_{kr}(t)$$

The desired outputs are then

$$C_{dj}(t) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(t,\tau)\left[M_k(t-\tau)+P_k(t-\tau)\right] d\tau \qquad \text{(B-1)}$$

Here a time variable weighting function $h(t,\tau)$ is defined as the response of a system at time t to a unit impulse applied at time $t-\tau$, i.e., τ time-units earlier.

The system errors are expressed by

$$\begin{aligned}\epsilon_j(t) &= C_j(t)-C_{jd}(t)\\ &= \sum_{k=1}^{n} \int_{o}^{t_o} W_{jk}(t,\tau)\left[M_k(t-\tau)+N_k(t-\tau)\right.\\ &\qquad \left.+ P_k(t-\tau)\right]\sum_{i=-\infty}^{\infty} \delta(t-\tau-iT)d\tau\\ &\quad - \sum_{k=1}^{n} \int_{-\infty}^{\infty} D_{jk}(t,\tau)\left[M_k(t-\tau)+P_k(t-\tau)\right] d\tau\end{aligned} \qquad \text{(B-2)}$$

For $\overline{\epsilon_j(t)} = 0$, the constraint equations are, with $t = hT+\Delta$,

$$\begin{aligned}&\sum_{q=o}^{N} p_{kr}(hT-qT)W_{jk}(hT+\Delta, qT+\Delta)\\ &\quad = \int_{-\infty}^{\infty} p_{kr}(t-\tau)D_{jk}(t,\tau)d\tau \qquad \text{(B-3)}\end{aligned}$$

$$k = 1, 2, \ldots n$$

$$r = 0, 1, \ldots m_k$$

where $W(t,\tau) = 0$ for $(N+1)T < \tau < 0$ and $t < \tau$ (or $h < q$).

If these constraints on $W(t,\tau)$ are satisfied the mean square errors are

$$\begin{aligned}\overline{\epsilon_j^2(t)} &= \overline{\epsilon_j^2(hT+\Delta)}\\ &= \sum_{k=1}^{n}\sum_{k'=1}^{n}\sum_{\ell=o}^{N}\sum_{q=o}^{N} W_{jk'}(hT+\Delta,\ell T+\Delta)\\ &\qquad \times W_{jk}(hT+\Delta,qT+\Delta)\phi_{k'k}(hT-\ell T,hT-qT)\\ &\quad - 2\sum_{k'=1}^{n}\sum_{\ell=o}^{N} W_{jk'}(hT+\Delta,\ell T+\Delta)\\ &\qquad \times B_{jk'}(hT+\Delta,\ell T+\Delta)\\ &\quad + \sum_{k=1}^{n}\sum_{k'=1}^{n}\int_{-\infty}^{\infty} D_{jk'}(t,\tau_2)d\tau_2\\ &\qquad \times \int_{-\infty}^{\infty} D_{jk}(t,\tau_1)\phi_{M_{k'},M_k}(t-\tau_2,t-\tau_1)d\tau_1\end{aligned} \qquad \text{(B-4)}$$

where

$$\phi_{k'k}(t,\tau) = \phi_{M_{k'},M_k}(t,\tau)+\phi_{M_{k'},N_k}(t,\tau) + \phi_{N_{k'},M_k}(t,\tau)+\phi_{N_{k'},N_k}(t,\tau)$$

$$\theta_{k'k}(t,\tau) = \phi_{M_{k'},M_k}(t,\tau)+\phi_{N_{k'},M_k}(t,\tau)$$

and

$$B_{jk'}(t,\tau_2) = \sum_{k=1}^{n}\int_{-\infty}^{\infty} D_{jk}(t,\tau_1)\theta_{k'k}(t-\tau_2,t-\tau_1)d\tau_1$$

By using the same minimization technique to the system ensemble mean square error, the simultaneous equations which the optimum weighting functions must be satisfied subjected to the constraints can be shown as

$$\sum_{k=1}^{n}\sum_{q=o}^{N}\phi_{k'k}(hT-\ell T,hT-qT)W_{jk}(hT+\Delta,qT+\Delta)$$

$$=B_{jk'}(hT+\Delta,\ell T+\Delta)-\sum_{r=o}^{m_{k'}}\lambda_{k'r}p_{k'r}(hT-\ell T) \qquad \text{(B-5)}$$

$$\text{for } k' = 1, 2, \ldots n$$

$$\ell = 0, 1, \cdots N$$

In order to use matrix algebra to solve Eqs. (B-3) and (B-5), a set of submatrices $\underline{H}$, $\underline{I}$, $\underline{J}$, $\underline{F}_j$, $\underline{G}_j$, $\underline{E}_j$ and $\underline{V}_j$ can also be defined. The major difference from the previous treatment is that the matrices now are a function of t (or h) as well as Δ. Except for the simplest problem, use of digital computer is advisable in obtaining the solution.

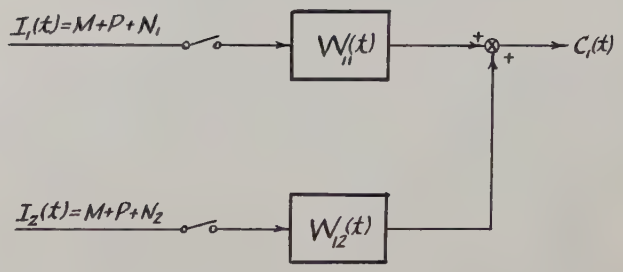


FIG 3 2+1 Poles Filter

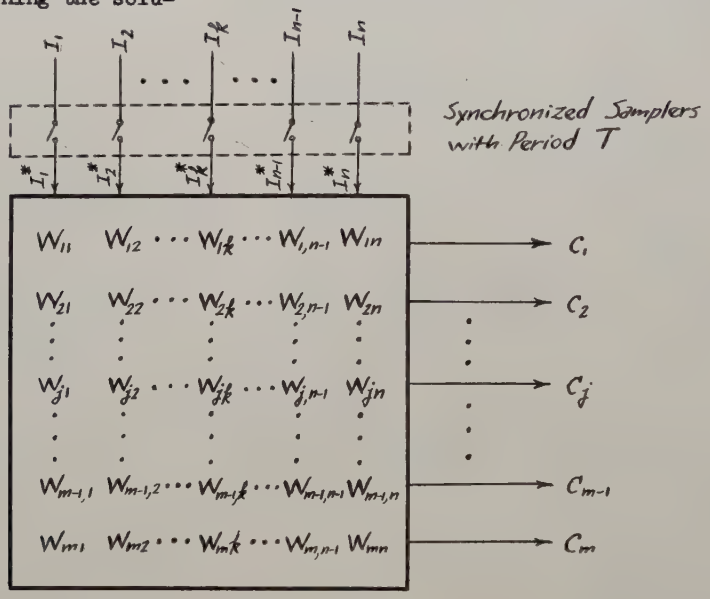


FIG 1 Block Diagram for n+m Poles Filter

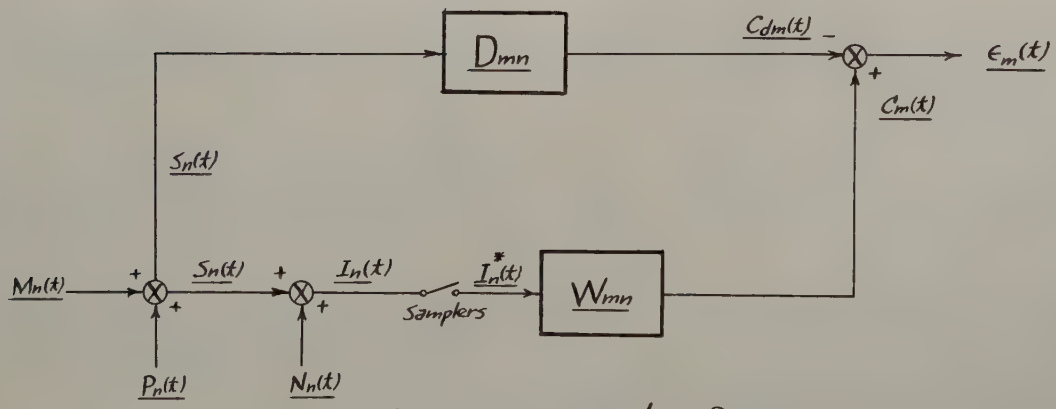


FIG 2 Error Generation Diagram

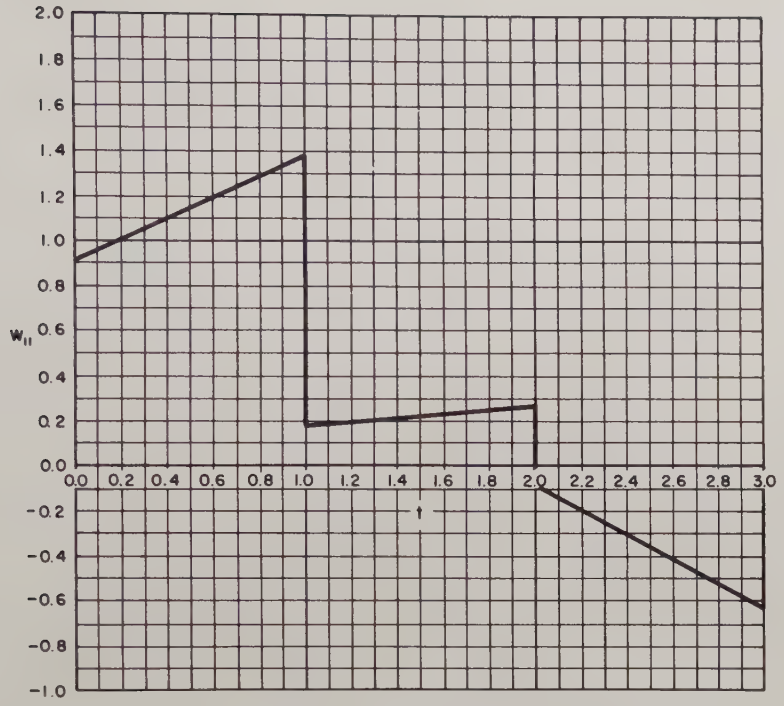


Fig. 4(a) Weighting function for the example.

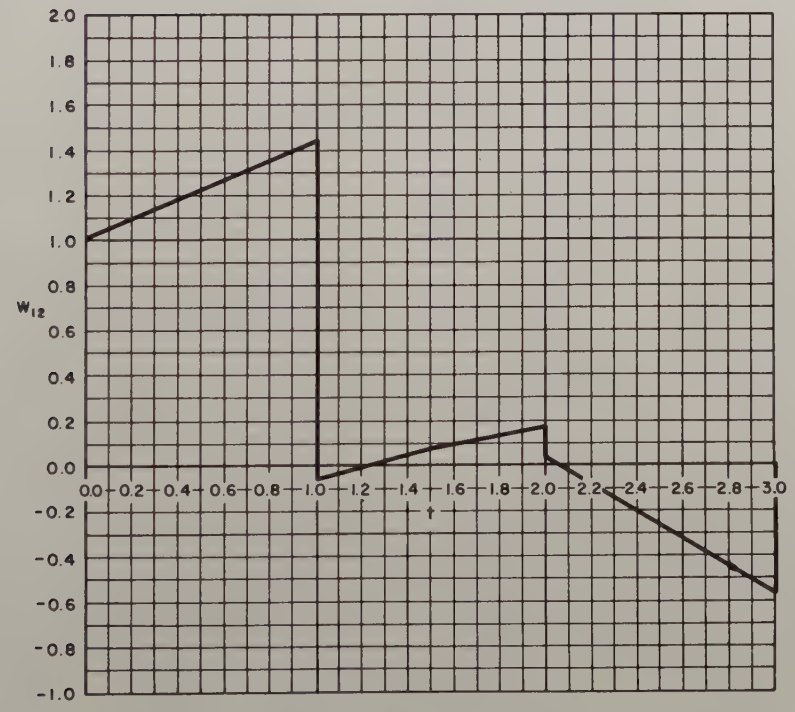


Fig. 4(b) Weighting function for the example.

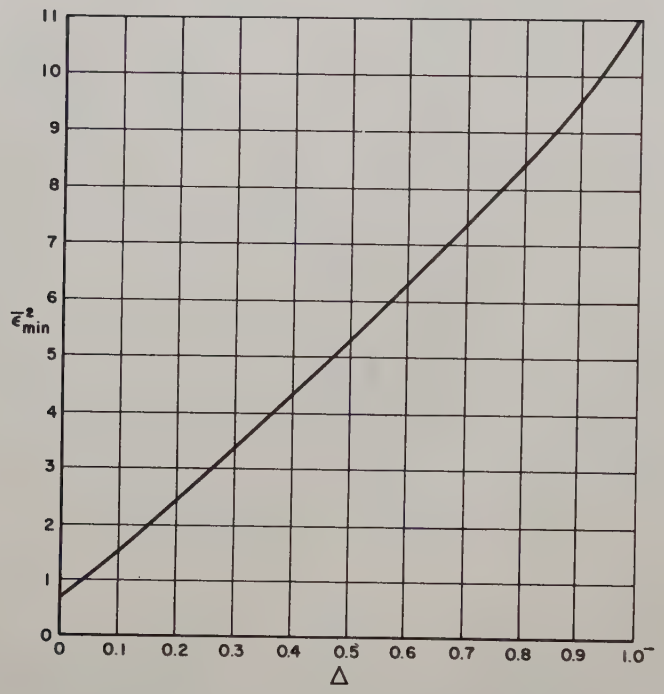


Fig. 5. Minimum mean square error.

DECOUPLING TECHNIQUES IN MULTILOOP CONTROL SYSTEMS

R. H. Loomis
Westinghouse Electric Corporation
Baltimore, Maryland

Summary

The stability and accuracy of a control system consisting of two interwoven loops are affected by interaction among all transfer functions and all inputs. This is undesirable, since the design of each transfer function must be compromised to account for the interaction effects.

In some cases, the two loops can be decoupled by the addition of an appropriate feedback path. This paper describes the application of this technique to a radar angle tracking loop.

The technique described has the following advantages:

1. The required matching takes place within the loop so that the effect of any mismatch is attenuated by the open loop gain.

2. For the example used, the output angular rate accuracy becomes independent of aircraft motion.

3. Subject to the rather minor restriction implied by item 1, the two loops become independent as far as stability is concerned.

Decoupling, which is described in this paper by illustrating its application to a particular system, is applicable to any complex control system. However, to the author's knowledge, no formal procedure for general application is available.

Introduction

Multiloop Control Systems

This paper outlines a technique whereby the several interwoven loops of a complex control system can be made to act, dynamically, as if they were independent and operating in cascade. This technique can be useful in improving system stability in cases where the transfer functions are fixed by other considerations. It is especially useful in cases where the controlled variable is not of principal concern (e.g., when output rate accuracy is of principal interest) and has been successfully applied to a system having this particular characteristic.

The control loop shown in block diagram form as figure 1 is typical of the class of loops to which decoupling can be usefully applied.

The outer or control loop serves to track the reference variable, R, while the inner or isolation loop acts to prevent system response to the disturbance function, U. However, since they are interwoven, both the control and isolation characteristics will be affected by both transfer functions. For example, the outer feedback path (through G_1) can act to improve the system isolation capability and decrease the phase margin.

Decoupling

The destabilizing effect of the outer loop can be counteracted by appropriate modification of the inner loop transfer function. However, if this approach is used, the added isolation brought about by the outer loop is lost as the design of each loop must be compromised to account for the effect of the other loop.

An alternate approach is to change the tracking configuration in order to prevent interaction between the loops. This can be done quite easily, although at a sacrifice in control "tightness." (That is, the output variable (C) will not follow the input variable (R) as well.)

In the example discussed later, the effect of decoupling is to make the system shown in figure 2 dynamically equivalent to that system shown in figure 3.

The use of a decoupling loop introduces problems of drift and component matching. The component matching problems are not significant, since the matching is done within the loop and, therefore, the effects of any mismatch are attenuated by loop gain. The drift problem, however, requires careful attention to component selection.

Conditions

Decoupling is not always a desirable approach to a stability problem, since the increased stability is obtained at the expense of loop "tightness." It should be used only after a

study of accuracy requirements has shown that the resultant loss of accuracy is acceptable.

The following conditions are examples of the factors which must be present to justify consideration of decoupling:

1. A system of the general type shown in figure 2 is to be used. More specifically, a complex control system is to be used.

2. The transfer functions cannot be designed to meet all requirements and still retain adequate overall phase margin.

3. Some degradation in output accuracy is acceptable.

A situation which meets these conditions is discussed under "Practical Use of Decoupling." The application of the technique is as described under "Application of Decoupling."

In addition to the above usage criteria, the application of decoupling places requirements on all components in the loops involved. Although the principal purpose of this paper is to present an analytical description of decoupling techniques, the successful application of these techniques is so heavily dependent on the use of appropriate components that a brief discussion of the component selection problem is included under "Component Matching" and "Component Drift."

Application of Decoupling

System Configuration

The control system shown as figure 2 is typical of the class to which decoupling can be applied effectively, and it can be used to illustrate the decoupling technique. For clarity, a specific open loop transfer function will be assumed. Figure 4 is an open loop gain frequency plot of the assumed function.

Original Loop Stability

In the assumed control system (figure 2), the outer loop can be regarded as a feedback path around the isolation loop for stability analysis. This has been done for a particular group of parameters, and the result is shown in figure 4.

Examination of the figure shows that the control loop feedback has two effects (other than its primary tracking function). First, it increases the isolation performance of the loop, and second, it decreases the overall phase margin. For the assumed parameters, the decrease in phase margin is such that the overall loop phase margin is reduced to an unacceptable value.

The magnitude of this reduction is calculated as follows:

$$H = s \tag{1}$$

Adding the control loop, the total feedback becomes:

$$H' = s + G_1 \tag{2}$$

Taking the ratio between the two feedbacks and using assumed parameter values give:

$$\frac{GH'}{GH} = \frac{s + G_1}{s} = \frac{G_1}{s}(1 + s/G_1) = \frac{20}{s}(1 + s/20) \tag{3}$$

The outer loop introduces an increase in slope below a frequency of 20 radians per second. This is illustrated in figure 4 where the solid lines show the stabilization loop alone and the dashed lines show the overall system. The destabilizing effect is apparent from the figure. For these parameters, the loss in phase margin is:

$$90° - \tan^{-1}\frac{33}{20} = 33° \tag{4}$$

This loss of 33 degrees reduces the overall phase margin to approximately 10 degrees, which is not a satisfactory value.

Original Loop Performance

The performance of the original system can be obtained by direct solution using the block diagram shown as figure 2. Thus

$$C = \frac{1}{1 + s/G_1 + s/G_1G_2}R + \frac{s/G_1G_2}{1 + s/G_1 + s/G_1G_2}U \tag{5}$$

The added isolation from G_1 is apparent from the equation. Figure 10 shows the response of this loop to a severe transient in U. From this figure, it is apparent that excellent positional isolation, as well as low phase margin, is afforded.

An obvious method of improving the phase margin is to add a compensating network to G_2. However, this approach would also eliminate the added isolation brought about by the outer loop and, in some cases, redesign of the loop would be severely limited. (For example, the quadratic indicated in figure 4 might be due to a mechanical resonance beyond the designers control.) This being the case and taking into consideration the conditions of a previous section, decoupling may

offer a more promising approach.

Decoupling Loop

The control system of figure 2 can be decoupled by adding a feedback path from the output of the integrator to the input to G_1. In general, this would be quite simple to mechanize. Figure 5 is a block diagram of the decoupled configuration.

Decoupled Loop Stability

The effect of decoupling on stability can best be examined by writing the control equations for the loop. This gives:

$$C' = \left(\frac{G_1/s}{1+G_1/s}\right)\left(\frac{G_2}{1+G_2}\right) R + \left(\frac{1}{1+G_2}\right) U \qquad (6)$$

Equation 6 can now be used to derive the dynamic equivalent of the decoupled configuration. Figure 6, a block diagram of this equivalent system, shows clearly that the effect on stability of loop interaction has been completely eliminated. The addition of the decoupling loop makes the resultant system act as if it were comprised of two independent loops operating in cascade. This has obvious advantages in design. Since the loops are independent as far as stability is concerned, a considerably wider choice of parameters is available to the designer. However, for the purpose of this paper, the transfer functions will be considered fixed and the effect of decoupling the original loops evaluated.

Decoupled Loop Performance

Equation 6 indicates the effect of decoupling on the accuracy and isolation capability of the system. If this equation is rewritten, it can be put in a form more suitable for comparison with equation 5 which is the equivalent relation for the original system.

$$C' = \frac{1}{1+s/G_1+s/G_1G_2+1/G_2} R + \frac{1/G_2(1+s/G_1)}{1+s/G_1+s/G_1G_2+1/G_2} U \qquad (7)$$

Repeating equation 5:

$$C = \frac{1}{1+s/G_1+s/G_1G_2} R + \frac{s/G_1G_2}{1+s/G_1+s/G_1G_2} U \qquad (5)$$

Comparison of these relations shows two differences:

1. A term, $1/G_2$, has appeared in the denominator of both terms; if G_2 is a high gain at tracking frequencies, this term will have little effect on the tracking response.

2. The numerator of the disturbance term has been changed by the addition of the term $1/G_2$; this gives a considerable reduction in the system isolation performance.

Since the first effect is usually negligible or can be made so, the second effect is of principal interest.

Figure 7 shows the response of both the original system (solid line) and the decoupled system (dashed line) to a severe transient to U. This figure shows the degradation in isolation performance which occurs. The transient used will be described under "Practical Use of Decoupling."

Review

In this section, decoupling was explained by direct application to a typical system. The use of decoupling had two principal effects:

1. It eliminated the dynamic interaction between inner and outer loops, thus easing the stability problem considerably.

2. It reduced the isolation capability of the system by a considerable amount.

In other words, the use of decoupling improved the stability characteristics at the expense of the isolation capability. In some applications, the loss of isolation capability is negligible. The next section describes a case (which uses the same system discussed here) in which the advantages of decoupling more than offset the reduced isolation performance.

Practical Use of Decoupling

System Configuration

The system discussed in "Application of Decoupling" (see figure 2) is typical of the type used for angle tracking in airborne radar. The previous section has shown the effects which decoupling will have on a system of this general type. This section indicates the effects which will occur when decoupling is applied to an airborne radar angle tracking loop of this same general type.

In an airborne radar, the angular rate accuracy is usually more critical than the angular position accuracy. This fact, combined with the stability considerations covered previously, makes decoupling a useful technique in the design of this type of loop.

The system shown previously as figure 2 is used in the following discussion. Its block diagram is redrawn as figure 8 to show the rate output.

Original Loop Stability and Performance

Since this loop is identical to that used in the discussion under "Application of Decoupling," the stability and performance characteristics are identical. However, the rate signal is now of primary interest. Equations 8 and 9 describe the position and rate performance.

$$C = \frac{1}{1+s/G_1+s/G_1G_2}R + \frac{s/G_1G_2}{1+s/G_1+s/G_1G_2}U \qquad (8)$$

$$V = \frac{s(1+1/G_2)}{1+s/G_1+s/G_1G_2}R - \frac{s/G_2}{1+s/G_1+s/G_1G_2}U \qquad (9)$$

Equation 8 is identical to equation 5 which has been discussed in the section entitled "Application of Decoupling."

Equation 9 which describes the rate performance of the loop shows several interesting features.

1. The terms s/G_2 will tend to amplify high frequency noise in the input.

2. The disturbance present in the positional term appears in the rate term multiplied by G_1.

This indicates that the rate term is much more sensitive to both disturbance, U, and noise in the input signal than is the positional signal. Since this term is used (after multiplication by a prediction factor which is usually greater than unity) to predict target future position in airborne fire control systems, the presence of these noise signals is undesirable.

These noise considerations, combined with the low stability of the system (using the same assumed transfer functions as before - see figure 4) indicate that this loop design is not satisfactory. In this instance, the use of a decoupling loop will give a considerable improvement in system performance.

Decoupled Loop Stability

Since the same configuration and transfer functions have been used here and in the section entitled "Application of Decoupling," the same change in stability will occur. The addition of the decoupling loop will make the overall system act, dynamically, as if two independent loops were operating in cascade. This is illustrated in figure 9.

The only stability requirement for the decoupled system is that G_2 and G_1/s, separately, be stable open loop transfer functions. The two transfer functions no longer interact and may be designed separately.

The transfer functions assumed here are such that the addition of the decoupling loop completely solves the stability problem. In addition, it permits a redesign of G_1 and G_2 to give practically any desired characteristic. However, this paper will not explore the possibilities in redesign of the transfer functions.

Decoupled Loop Performance

The rate performance of the angle tracking system has been improved by the decoupling loop while the positional performance has been degraded. This is indicated by the response equations.

$$C' = \frac{1}{(1+1/G_2)(1+s/G_1)}R + \frac{1/G_2}{1+1/G_2}U \qquad (10)$$

$$V' = \frac{s}{1+s/G_1}R \qquad (11)$$

The effect of platform disturbance has been completely eliminated from the rate term but its effect on the position term has been increased by a large factor (see figures 10 and 11).

Figures 7, 10, and 11 were obtained by analog simulation of a complex angle tracking control system. The disturbance was adjusted to be equivalent to a large rate disturbance. To be more precise, the disturbance used was equivalent to that seen by an airborne tracking radar with a lead angle of ≈ 20 degrees, when the aircraft rolls at 100 degrees per second.

Review

"Practical Use of Decoupling" has described the application of the type of feedback compensation we have termed decoupling to a simple loop. The effects obtained were, of course, the same as were indicated in "Application of Decoupling." However, several additional considerations became apparent.

1. The controlled variable was not of principal interest, and, although the position control was degraded, a complete cancellation of the disturbance effect on V was achieved.

2. Since the decoupling makes the loops act, dynamically, as if they were independent and operating in cascade, the two transfer functions can be designed independently.

3. The degradation in output positional accuracy is considerable and care must be taken to ensure that this does not become excessive.

Component Matching

General

The preceding discussion is quite restrictive in that the transfer functions used have been oversimplified in order to describe the technique. Practically speaking, there must be transfer functions in both feedback paths of figure 5 to represent the physical elements necessary to convert mechanical motions to electrical signals, transmit the signals, and so forth. These elements are usually not represented in simple systems, since their effects can, generally, be absorbed in other transfer functions without loss of realism. However, in the case at hand, absorbing the effects of these elements in other transfer functions or omitting them from consideration implies severe restrictions on the physical elements involved in the feedback paths. That is, the extent to which a given loop can be decoupled and the utility of the resultant loop are limited by the characteristics of individual components and by the extent to which components can be matched.

Feedback Transfer Functions

Figure 5 can be redrawn, as illustrated in figure 12, to include the effects of the feedback transfer functions. We have shown the decoupling feedback as going to the internal summing point in order to represent more closely the mechanization which would actually be used in an airborne tracking loop.

This tracking configuration gives the following response equations:

$$C' = \frac{G_1G_2G_3}{1+G_1G_2G_3G_6+G_2G_3G_4+G_1G_3G_5}R + \frac{1+G_1G_3G_5}{1+G_1G_2G_3G_6+G_2G_3G_4+G_1G_3G_5}U \quad (12)$$

$$V' = \frac{G_1(1+G_2G_3G_4)}{1+G_1G_2G_3G_6+G_2G_3G_4+G_1G_3G_5}R + \frac{G_1(G_3G_4G_5) - G_1G_6}{1+G_1G_2G_3G_6+G_2G_3G_4+G_1G_3G_5}U \quad (13)$$

These equations can be rewritten in a form suitable for direct comparison to equations 10 and 11 as follows:

$$C' = \frac{1}{G_6+1/G_1G_2G_3+G_4/G_1+G_5/G_2}R + \frac{1/G_2\left[G_5+1/G_1G_3\right]}{G_6+1/G_1G_2G_3+G_4/G_1+G_5/G_2}U \quad (14)$$

$$V' = \frac{G_4+1/G_2G_3}{G_6+1/G_1G_2G_3+G_4/G_1+G_5/G_2}R + \frac{1/G_2\left[G_4G_5-G_6/G_3\right]}{G_6+1/G_1G_2G_3+G_4/G_1+G_5/G_2}U \quad (15)$$

If equations 14 and 15 are compared to equations 10 and 11, several differences become apparent, and the condition which the individual transfer functions must meet, if successful decoupling is to be accomplished, is readily derived.

Requirements for Decoupling

Consider, first, the terms which provide the cancellation of disturbance, U, in the rate term, V. These terms appear as the numerator of the transfer function for U in equation 15.

$$f_V(U) = \frac{1/G_2\left[G_4G_5 - G_6/G_3\right]}{G_6+1/G_1G_2G_3+G_4/G_1+G_5/G_2}U \quad (16)$$

If V is to be unaffected by U, we have the requirement

$$G_4G_5 = G_6/G_3 \quad (17)$$

If the loop is to be decoupled, the characteristic function must be separable into two factors, one factor independent of G_2 and the other independent of G_3. This will occur provided:

$$G_3G_4G_5 = G_6 \quad (18)$$

This is exactly the same requirement as was obtained for the elimination of U from the rate signal.

The function, G_6, is not available to the designer of an airborne angle tracking loop, since it is operating on the space positional feedback. For purposes of discussion, let us assume that G_5 is adjusted to equal the obtained value of G_6 and that G_3 is adjusted to equal $1/G_4$. (Note that here we are artificially introducing a double matching problem; only one matching, given by equation 18, is basically required.)

$$C' = \frac{1}{(G_5+1/G_1G_3)(1+1/G_2)}R + \frac{1/G_2}{(1+1/G_2)}U \qquad (19)$$

$$V' = \frac{1/G_3}{G_5+1/G_1G_3}R \qquad (20)$$

If these two requirements are imposed, the decoupled loop has the dynamic equivalent shown as figure 13.

The equations from which the diagram is derived follow directly from the imposition of the above conditions on equations 14 and 15.

Figure 14 illustrates the dynamic equivalent which is obtained if $G_3G_4G_5$ is taken equal to G_6 without the further assumption used to obtain figure 13.

The equations from which figure 14 is derived are as follows:

$$C' = \frac{1}{\left[G_5+1/G_1G_3\right]\left[G_6/G_5+1/G_2\right]}R + \frac{1/G_2}{G_6/G_5+1/G_2}U \qquad (21)$$

$$V' = \frac{1/G_3}{G_5+1/G_1G_3}R + \frac{1/G_2 \cdot 0 \cdot U}{G_6/G_5+1/G_2} \qquad (22)$$

Review

As indicated by equations 18 through 21, the decoupling of the loop is strongly affected by the extent to which certain transfer functions are matched. This matching problem is not, however, as severe for the decoupling approach as it normally is for matching filters, feed-forward, and conditional feedback techniques, since, in this case, the matching is done within the control loop, and the effect of mismatch is attenuated by the loop gain (see equations 14 and 15).

The loops as shown in figure 14 are not completely decoupled, since both loops involve the function G_5. Note, however, that the loops can still be designed independently provided that G_5 is matched to G_6. Even if these two components are not matched, considerably more latitude is available in the design of G_1 and G_2 than is available without the use of a decoupling loop*.

*Throughout this paper, it is assumed that the transfer functions completely describe the loop characteristics, including loading effects.

Component Drift

General

Precise noise and drift free components must be used in the decoupling feedback path. The following paragraphs discuss the effect of drift in various components and serve to provide a mathematical basis for this statement.

Figure 15 illustrates the drifts to be discussed.

Transfer Functions

Direct solution of the block diagram shown as figure 15 gives the following response equations:

$$G_1G_2G_3R - G_1G_2G_3G_5D_1 - G_2G_3G_4D_2 - G_1G_2G_3G_6D_3 + \left[1+G_1G_3G_5\right]U = \left[1+G_1G_2G_3G_6 + G_1G_3G_5 + G_2G_3G_4\right]C' \qquad (23)$$

$$G_1(1+G_2G_3G_4)R - G_1G_5(1+G_2G_3G_4)D_1 + G_1G_3G_4(G_5+G_2G_6)D_2 - G_1G_6(1+G_2G_3G_4)D_3 + G_1(G_3G_4G_5-G_6)U = (1+G_1G_2G_3G_6+G_1G_3G_5+G_2G_3G_4)V' \qquad (24)$$

If we now assume that the decoupling circuit is designed properly (i. e., that $G_3G_4G_5 = G_6$), equations 23 and 24 reduce to the following form.

$$C' = \frac{1}{D}R - \frac{G_5D_1}{D} - \frac{(G_4/G_1)D_2}{D} - \frac{G_6D_3}{D} + \frac{(1/G_2)U}{G_6/G_5+1/G_2} \qquad (25)$$

$$V' = \frac{(1/G_3)R}{D'} - \frac{(G_5/G_3)D_1}{D'} + \frac{G_4G_5D_2}{D'} - \frac{(G_6/G_3)D_3}{D'} + \frac{(1/G_2)\,0\,U}{D} \qquad (26)$$

where: $D = (G_5+1/G_1G_3)(G_6/G_5+1/G_2)$

$$D' = G_5 + 1/G_1G_3$$

For comparison, the equations describing drift performance for the original loop are:

$$\left[1+G_1G_2G_3G_6+G_2G_3G_4\right]C = G_1G_2G_3R + U - G_2G_3G_4D_2 - G_1G_2G_3G_6D_3 \qquad (27)$$

$$\left[1+G_1G_2G_3G_6+G_2G_3G_4\right]U = G_1(1+G_2G_3G_4)R - G_1G_6U + G_1G_2G_3G_4G_6D_2 - G_1G_6\left[1+G_2G_3G_4\right]D_3 \qquad (28)$$

These can be put in a form more suitable for comparison to equations 25 and 26 as follows:

$$C = \frac{R}{F} + \frac{1/G_1G_2G_3}{F} U - \frac{G_4/G_1}{F} D_2 - \frac{G_6 D_3}{F} \qquad (29)$$

$$V = \frac{\left[G_4 + 1/G_2G_3\right] R}{F} - \frac{\left[G_6/G_2G_3\right] U}{F} + \frac{G_4G_6 D_2}{F} - \frac{\left[\frac{G_6}{G_2G_3}\right] + G_4G_6 D_3}{F} \qquad (30)$$

where: $F = (1/G_1G_2G_3) + G_4/G_1 + G_6$

Comparison

Table 1 summarizes the performance of both the decoupled and the original loops.

Table 1 is not of immediate assistance in obtaining a comparison between the original and the decoupled loops, since, in the general case, a wide range of parameter values could be used for each of the transfer functions. If the following assumptions are made, a more obvious comparison, which is typical of airborne angle tracking loops, is obtained.

The result of applying these assumptions to the relations of table 1 is shown as table 3.

Review

For the angle tracking loop defined by the assumptions of table 2, the addition of the decoupling network leaves the drift characteristics essentially unchanged (G_2 is generally a high gain), except that it provides an additional loop within which drift can occur.

The effect of drift on the more general configuration depends upon the characteristics of the transfer functions in the manner shown in table 1.

TABLE 1

COMPARISON BETWEEN ORIGINAL AND DECOUPLED LOOP RESPONSE EQUATIONS - GENERAL

INPUT	C RESPONSE ORIGINAL	C'RESPONSE DECOUPLED	V RESPONSE ORIGINAL	V' RESPONSE DECOUPLED
R	$1/F^*$	$1/D^{**}$	$\frac{1+G_2G_3G_4}{G_2G_3F}$	$\frac{1}{G_3D'^{***}}$
U	$\frac{1}{G_1G_2G_3F^*}$	$\frac{G_5}{G_2G_6+G_5}$	$-\frac{G_6}{G_2G_3F^*}$	———
D_1	———	$-\frac{G_5}{D^{**}}$	———	$\frac{G_5}{G_3D'^{***}}$
D_2	$-\frac{G_4}{G_1F^*}$	$-\frac{G_4}{G_1D^{**}}$	$\frac{G_4G_6}{F^*}$	$\frac{G_4G_5}{D'^{***}}$
D_3	$-\frac{G_6}{F^*}$	$-\frac{G_6}{D^{**}}$	$\frac{G_6(1+G_2G_3G_4)}{G_2G_3F^*}$	$-\frac{G_6}{G_3D'^{***}}$

* $F = 1/G_1G_2G_3 + G_6 + G_4/G_1$

** $D = \left[G_5 + 1/G_1G_3\right]\left[G_6/G_5 + 1/G_2\right]$

*** $D' = G_5 + 1/G_1G_3$

TABLE 2
ASSUMPTIONS

AFFECTED LOOP(S)	ASSUMPTIONS
Decoupled	$G_5 = G_3G_4G_5 = G_6 = 1$
Original	$G_5 = G_3G_4 = G_6 = 1$
Both	$G_1G_2 \longrightarrow \alpha$

Bibliography

1. Boksenbom, A. S. and Hood, H., "General Algebraic Method Applied to Control Analysis of Complex Engine Types," NACA Report 980, 1950.

2. Honnell, P. M., "The Transmission Matrix Stability Criterion," The Matrix and Tensor Quarterly, Vol. II, No. 1, September 1951, p. 8.

3. Honnell, P. M., "The Generalized Transmission Matrix Stability Criterion," AIEE Transactions, Vol. 70, pt. 1, 1951, p. 292.

4. Golomb, N. and Usdin, E., "A Theory of Multi-Dimensional Servo Analysis," Jour. Franklin Institute, Vol. 253, p. 29, 1952.

5. Boksenbom, A. S. and Hoppler, H., "Optimum Controllers for Linear Closed-Loop Systems," NACA Technical Note 2939, April 1953.

6. Cruickshank, A. J. O., "Matrix Formulation of Control System Equations," The Matrix and Tensor Quarterly, Vol. 5, No. 3, p. 76, 1955.

TABLE 3

COMPARISON BETWEEN ORIGINAL AND DECOUPLED LOOP RESPONSE EQUATIONS - SPECIAL CASE

INPUT	C RESPONSE ORIGINAL	C' RESPONSE DECOUPLED	V RESPONSE ORIGINAL	V' RESPONSE DECOUPLED
R	$1/F_A^*$	$1/D_A^{**}$	$\frac{1}{G_3}\left[\frac{1+G_2}{G_2F_A^*}\right]$	$\frac{1}{G_3D'_A}$
U	———	$\frac{1}{1+G_2}$	$-\frac{G_4}{G_2F_A^*}$	———
D_1	———	$-1/D_A^{**}$	———	$-\frac{1}{G_3D'_A}$
D_2	$-\frac{G_4}{G_1F_A^*}$	$-\frac{G_4}{G_1D_A^{**}}$	$\frac{G_4}{F_A^*}$	$\frac{G_4}{D'_A}$
D_3	$-1/F_A^*$	$-1/D_A^{**}$	$\frac{G_4(1+G_2)}{G_2F_A^*}$	$\frac{G_4}{D'_A}$

* $F_A = 1 + G_4/G_1 = D'_A$

** $D_A = \left[1 + G_4/G_1\right]\left[1 + 1/G_2\right]$

7. Povejsil, D. J. and Fuchs, A. M., "A Method for the Preliminary Synthesis of a Complex Multiple-Loop Control System," AIEE Applications and Industry, No. 19, p. 129, 1955.

8. Pack, G. J. and Phillips, Jr., W. E., "Analog Study of Interacting and Non-Interacting Multiple-Loop Control Systems for Turbojet Engines," NACA Report 1212, 1955 (Supercedes NACA Note 3112, 1954).

9. Kavanagh, R. T., "The Application of Matrix Methods to Multi-Variable Control Systems," Jour. Franklin Institute, Vol. 262, No. 5, p. 349, November 1956.

10. Freeman, H., "A Synthesis for Multipole Control Systems," AIEE Transactions, Vol. 76, Pt. II, 1957, p. 28.

11. Ragazzini, J. R., and Sarachik, P., "A 2-Dimensional Feedback Control System," AIEE Transactions, Vol. 76, Pt. II, p. 55.

12. Kavanagh, R. J., "Noninteracting Controls in Linear Multivariable Systems," AIEE Transactions, Vol. 76, Pt. II, 1957, p. 95.

13. Honnell, P. M., and Wolffenstein, D., "Transfer Matrix Stability Criterion Applied to an Amplidyne Servo Network," AIEE Transactions, Vol. 76, Pt. II, 1957, p. 143.

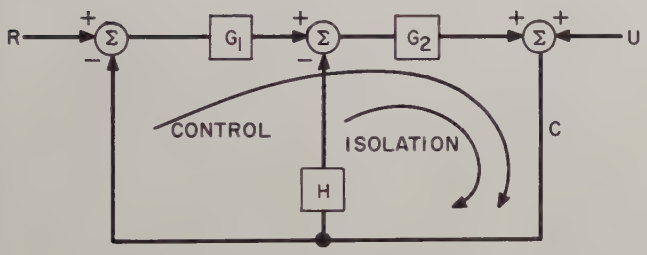


Fig. 1. Typical multiloop control system.

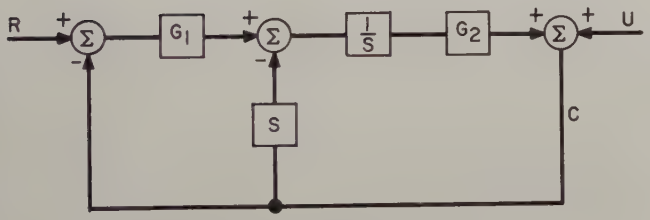


Fig. 2. Angle tracking loop.

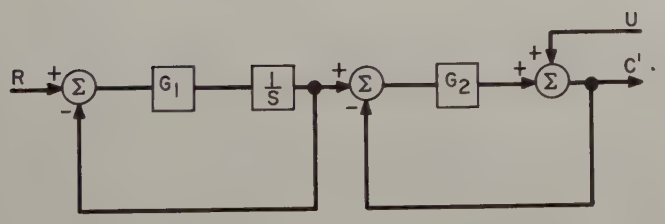


Fig. 3. Decoupled angle track loop dynamic equivalent.

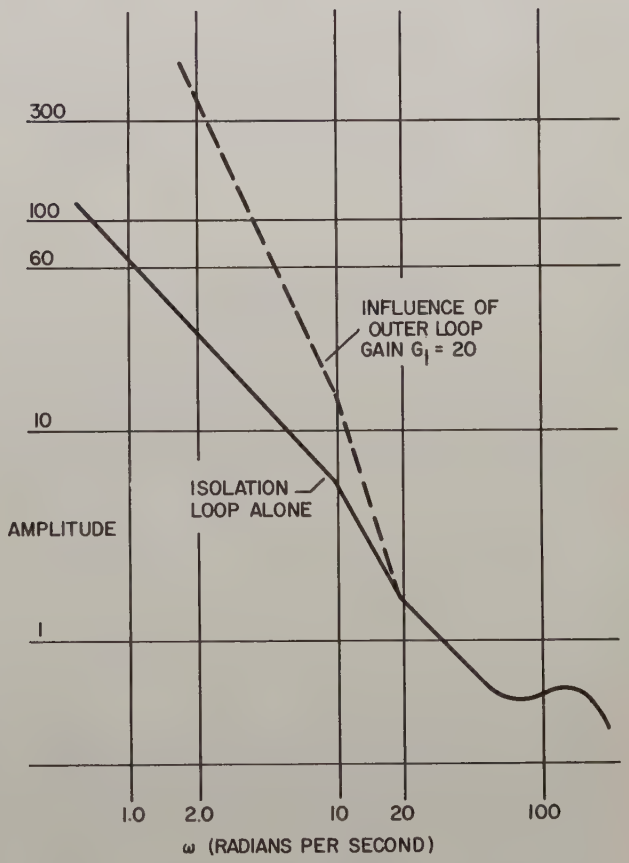


Fig. 4. Approximate gain-frequency curve for assumed tracking loop.

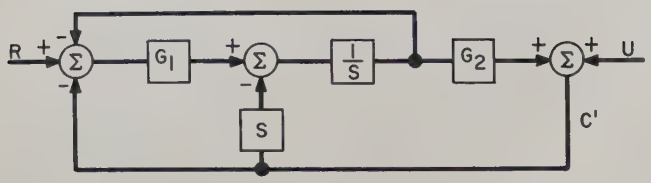


Fig. 5. Angle tracking loop, decoupled.

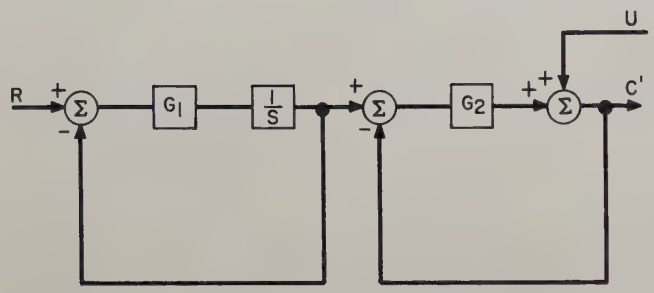


Fig. 6. Decoupled system—dynamic equivalent.

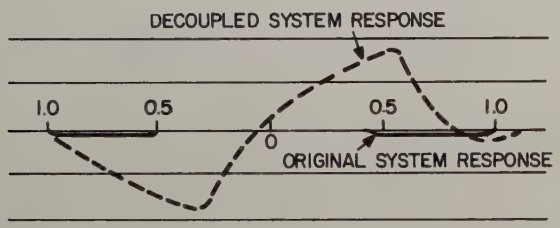


Fig. 7. Response to transient disturbance.

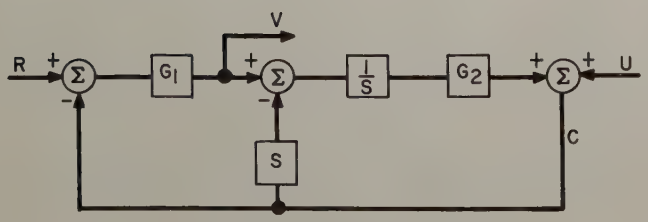


Fig. 8. Airborne radar angle tracking loop.

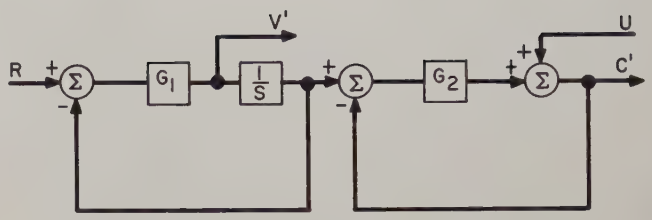


Fig. 9. Decoupled airborne angle tracking loop dynamic equivalent.

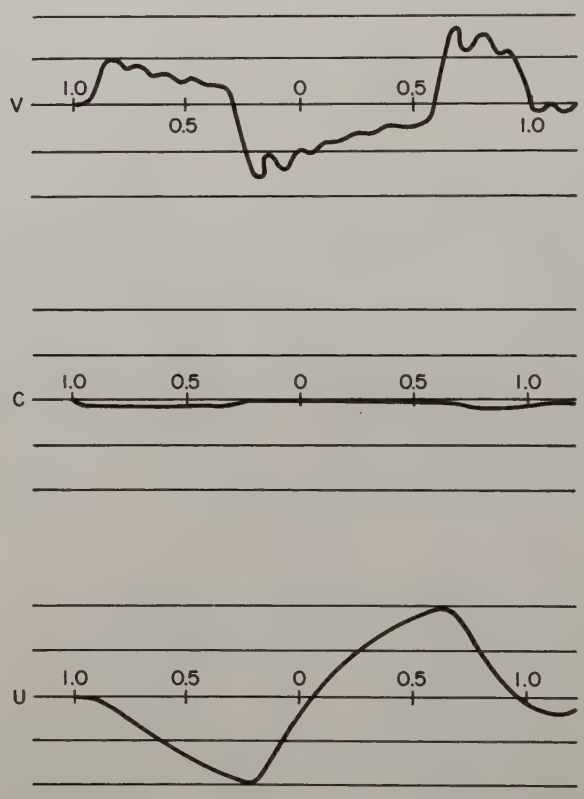


Fig. 10. Response of original system to a transient disturbance.

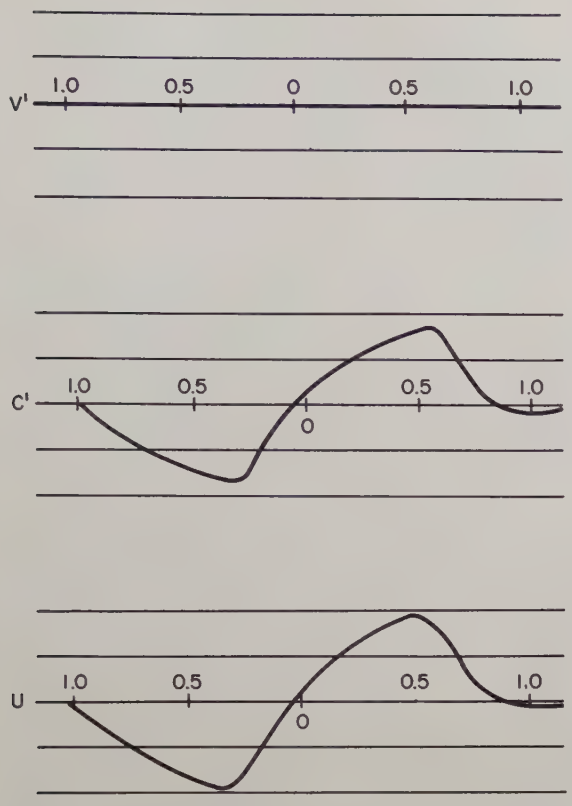


Fig. 11. Response of decoupled system to a transient disturbance.

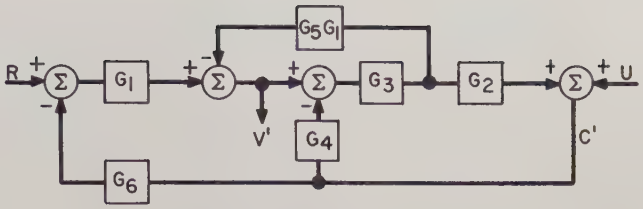


Fig. 12. Angle tracking loop, decoupled.

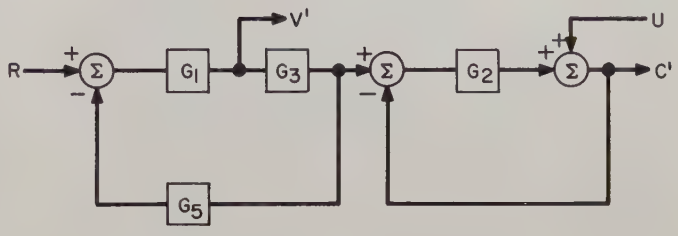


Fig. 13. Decoupled angle tracking dynamic equivalent.

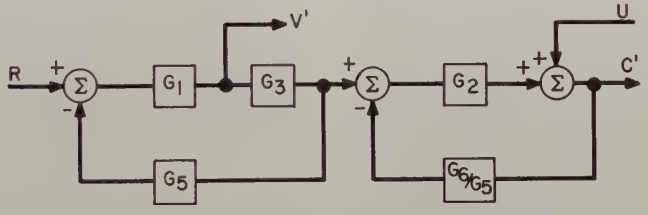


Fig. 14. Decoupled angle tracking loop.

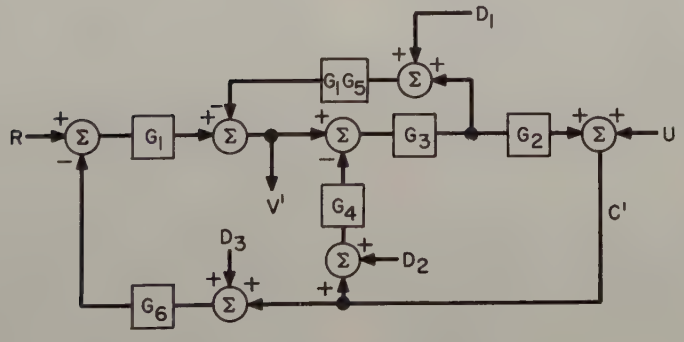


Fig. 15. Block diagram including drifts.

OPTIMUM COMPENSATION OF A POSITION SERVO WITH A MAGNETIC CLUTCH ACTUATOR

Ronald J. Hruby
Northrop Corporation, Norair Division
Hawthorne, California

Summary

In the past decade improvements have been made in the design of dry powder magnetic clutches which now make them competitive with hydraulic devices for missile servo actuator applications. Some of the inherent advantages of magnetic clutch actuators are: (a) increased control system reliability, (b) simplified equipment requirements, (c) properties remain constant with usage, (d) environmental requirements are quite compatible with an orbiting vehicle. The dynamic servo response of the magnetic clutch servo actuator is required in order to incorporate these advantages in missile and space control system design.

This paper presents a synthesis procedure for the optimum design of a position servo based upon the closed-loop transient response to a step input. The transient response of the position servo is described by four independent parameters. These parameters are the rise time, damping ratio, undamped natural frequency, and the steady-state gain. The synthesis procedure is based upon a theoretical model of a dry powder magnetic clutch, which is an air-gap, iron-care transformer with a one-turn secondary. The air gap is filled with dry ferrite particles. To provide simple analytic functions, a definition for rise time is developed which corresponds to the first crossing of unity gain to a high order of accuracy. The theoretical model of the magnetic clutch has two, first-order time lags. These two time lags are derived from the properties of the theoretical clutch model. They arise from the inductive properties of the clutch excitation coil and the clutch rotor induced eddy current, which is the current in the equivalent transformer single-turn secondary. The compensation for the excitation coil time lag consists of excitation current feedback, which decreases the effective transfer function time constant. The induced rotor eddy current time lag is optimally compensated by the proper selection of feedback gains as outlined in the synthesis procedure.

The position servo drives an undamped inertial load. This is characteristic of a control system for an outer space vehicle or a guided missile which utilizes swivel rocket nozzles for control. It is also a first-order approximation to an airframe control servo utilizing neutral aerodynamic surfaces with low damping. The position servo characteristics are provided by a position feedback branch, and the basic servo stability is provided by a tachometer feedback. The theoretical model results in a third-order system which cannot be unstable. Practical magnetic clutches differ from the theoretical model primarily because they have ferromagnetic hysteresis. At high frequencies, hysteresis appears as a constant time delay which makes servo instability possible. An additional constraint to the synthesis procedure is given which guarantees a stable, closed-loop position servo with the desired performance characteristics.

The entire analysis in this paper is based upon the application of Root-Locus analytical methods to extract the theoretical roots of the third-order system describing the magnetic clutch actuator position servo.

Introduction

This paper presents a design philosophy which differs from the modern trend in servo analysis. A mathematical model of the position servo is formulated, which allows the exact synthesis of the closed-loop servo, based entirely upon the desired performance characteristics rather than the stability requirements of gain and phase margin. The mathematical model is based upon the simplification of the transfer function of the magnetic clutch by using an equivalent model of a transformer with a one-turn secondary. The result of this design procedure is an analytic solution of the third-order system for each of the system parameters, such as feedback and forward branch gains as a function of the ideal performance characteristics. The final step in the design is a stability analysis.

Driving Amplifiers and Magnetic Clutch Unit[1,2]

Magnetic clutches used in missile servo actuators consist of an excitation coil which creates a changing magnetic field in a ferrite particle clutch gap. The existence of the magnetic field allows the transfer of torque through the clutch to the output shaft.

The basic electrical-mechanical configuration of the amplifier-magnetic clutch unit is shown in Figure 1. An a-c or d-c motor is the torque driving unit. It is mechanically coupled to the torque input spur gears on each clutch. The output torque shaft of each clutch is mechanically coupled to the servo output shaft. The clutches rotate in opposite directions to each other. Clutch #1 provides counterclockwise torque and clutch #2 provides clockwise torque. The two driving amplifiers each consist of a two-input difference amplifier driving a pentode class-B output stage. The output voltage for each amplifier varies from a small quiescent value to a maximum positive value. The input difference amplifiers are connected so that phase

inversion is accomplished. When a positive voltage is applied to the input of the unit, the current in clutch #2 approaches zero and the current in clutch #1 increases so that the output shaft has a counterclockwise torque. If a negative voltage is applied to the input of the unit, the current in clutch #1 approaches zero and an increasing current flows through clutch #2 which provides a clockwise torque to the output shaft. The amplifier clutch unit operates as a class-B push-pull system. With zero input voltage, a quiescent current is established in each clutch coil which semi-locks the output shaft position at zero output torque. The torque available at the output shaft is due to the drive motor. Changes in either the motor or load characteristics are in no way reflected back to the clutch currents via the clutch body.

The voltage developed across the resistor (R) is fed back to the input difference amplifier for compensating the excitation coil time lag.

The problem of inductive loading is severe enough that the use of pentodes to make a voltage-to-current amplifier without feedback in place of a voltage-to-voltage amplifier with feedback is not adequate because the self-inductance of the clutch coils will reduce the plate voltage until the pentodes operate in a quasi-triode mode during current transients.

Analysis and Servo Compensation of Clutch Coil

Figure 2a shows the equivalent clutch electrical circuit and equation 1 shows the resulting transfer function for (i/e_1).

$$\frac{i}{e_1} = \frac{K_o}{R_L + LP} = \frac{K'_o}{1 + \tau_L P} \qquad (1)$$

where:

R_L = internal coil resistance

K_o = gain of voltage amplifier

L = excitation coil self-inductance

$$\tau_L = \frac{L}{R_L}$$

$$K'_o = \frac{K_o}{R_L}$$

Figure 2b shows the method of decreasing the excitation coil time constant (τ_L) to any degree by feeding back a voltage porportional to excitation current. The resulting transfer function relating the excitation current (i) to the input voltage (e_1) is given in equation (2). Both Figures 2a and 2b are equivalent circuits for the hardware system shown in Figure 1, without and with coil time lag compensation.

$$\frac{i}{e_1} = \frac{K_o}{R_L + R[1+K_o h]} \times \frac{1}{1 + \frac{LP}{R_L + R[1+K_o h]}} \qquad (2)$$

$$= \frac{K_1}{1 + \tau_a P}$$

where:

$$\tau_a = \frac{L}{R_L + R[1+K_o h]}$$

$$K_1 = \frac{K_o}{R_L + R[1+K_o h]}$$

h = gain of current feedback network

R = excitation current measuring resistor

If the feedback term $(K_o h)$ is large, the time constant (τ_a) for the coil and amplifier is small. There is a restriction that R be small to minimize the input power requirements. By adjusting K_o, h and R the value of (τ_a) can be made as small as desired. The feedback network gain (h) will be less than 1, and the network is required to provide a d-c path from the output side of (R) to the grid of the input difference amplifier, as well as to provide proper biasing of the input amplifier grid.

Derivation of Rotor Eddy Current Time Lag [2]

The second first-order time delay (τ_c) arises from the induced secondary current in the clutch rotor, which makes the entire unit act like a poor transformer. The rotor has conducting elements by the nature of its structure, and the elements are equivalent to a single-turn secondary. The interaction of the driving current and induced secondary current produces the experimentally measured flux time lag.

Neglecting hysteresis, viscosity, bearing frictions, windage and drive motor transients, and assuming the drive motor is ideally compounded, the relationship between the magnetic flux and the exciting coil current is given by equations (3) and (4). It is a very good approximation to assume that the shaft output torque is directly proportional to the magnetic flux. This approximation is derived by neglecting the drive motor loading transients and using a push-pull clutch arrangement.

Equations (3) and (4) are derived by treating the rotor as a single-turn secondary with an equivalent winding resistance (R_S) and summing all the voltages and magnetomotive forces separately.

Sum of all voltage sources in the equivalent secondary:

$$i_S R_S + 10^{-8} \frac{d\phi}{dt} = 0 \qquad (3)$$

Sum of all magnetomotive forces in the magnetic circuit:

$$Ni + i_S = \frac{\phi \ell_1}{0.4\pi\mu A_g} \qquad (4)$$

where: (cgs - electromagnetic units - unrationalized)

i = coil exciting current (amperes)

i_S = single-turn secondary current (amperes)

R_S = single-turn secondary equivalent resistance (ohms)

ℓ_1 = mean flux path length thru total clutch gap (centimeters)

A_g = magnetic circuit mean cross-sectional area of clutch gap (sq. cm.)

μ = permeability of clutch gap

N = number of primary turns (turns)

ν = flux to torque coefficient (ft. lbs./ Maxwell)

ϕ = magnetic flux (Maxwells)

All magnetomotive forces are developed in the clutch gap, since the permeability of the ferrite core is much greater than the permeability of the clutch gap. The permeability of the rotor is taken to be the same as the gap energizing ferrite core.[3]

The shaft output torque is given by:

$$T = \nu\phi \qquad (5)$$

Combining equations (3), (4) and (5) gives the results exhibited in equation (6), which shows the transfer function relating output torque to coil current (T/i), and the first-order flux time lag (τ_c).

$$\frac{T}{i} = \frac{0.4\pi\mu NA_g\nu}{\ell_1} \times \frac{1}{1 + \frac{0.4\pi\mu A_g}{\ell_1 R_S}10^8 p} = \frac{K_2}{1 + \tau_c p} \qquad (6)$$

where:

$$\tau_c = \frac{0.4\pi\mu A_g \cdot 10^{-8}}{\ell_1 R_S}$$

$$K_2 = \frac{0.4\pi\mu N A_g \nu}{\ell_1}$$

VACUUM: $\mu = 1.0$

CLUTCH GAP: $\mu \sim 50$ TO 500

FERRITE CORE: $\mu \sim 10^3$ TO 10^4

The parameter (ν) depends upon the driving motor torque characteristics and the clutch-motor gear linkage as well as the shear properties of the iron particles. (ν) would have to be evaluated experimentally. The flux time lag is proportional to the clutch area but inversely proportional to the rotor resistance (R_S).

The overall transfer function relating input voltage (e_1) to output torque T is given in equation (7). This transfer function includes compensation for (τ_L).

$$\frac{T}{e_1} = \frac{K_1 K_2}{[1 + \tau_a p][1 + \tau_c p]} \qquad (7)$$

Optimum Synthesis Procedure

In the preceding analysis the derivation and compensation of (τ_L) was developed. Because the value of (τ_a) can be made very small by adjusting K_o, h and R, it will be set equal to zero for the closed-loop synthesis. This procedure is justified because the complex closed-loop roots are relatively unaffected by open-loop poles which are an order of magnitude greater than (τ_c).

An optimum synthesis procedure must start with a set of performance parameters and provide analytic functions which compute the values of the various elements that constitute the closed-loop servo. These performance parameters must be related to some type of exciting function which has been standardized by usage. The synthesis developed in this paper is based upon a unit step exciting function and four performance parameters which are the natural consequence of the unit step driving a finite servo.

The damping ratio (ζ), the undamped natural (ω_n), the rise time (t_S), and the steady-state response ($e_{o\,s.s.}$) are the four performance parameters used in the synthesis procedure. The damping ratio (ζ), the undamped natural frequency (ω_n), and the steady-state response ($e_{o\,s.s.}$) are universally used performance parameters, but the rise time (t_S) does not have a universally accepted definition. In this paper the rise time will be defined as the time required for the sinisoidal term in the time domain to reach its first zero. For minimal values of (t_S), the output of the servo lies between 95% and 100% of its steady-state value at its first crossing for values of (ζ) between (.05) and (.95). Each of these performance parameters is shown perspectively in Figure 3. The definition for rise time (t_S) was chosen because the synthesis equations are closed and analytic with respect to the design parameters.

The complete closed-loop servo is composed of a forward branch which consists of an amplifier and clutch unit as shown in Figure 1, an

inertial load (J), and both position and tachometer feedback branches which have gains (a_1) and (a_2), respectively. The complete servo is shown in Figure 4. The transfer function relating the output position (e_o) to the input signal (e_i) is given in equation (8), which includes the preceeding assumptions.

$$\frac{e_o(p)}{e_i(p)} = \frac{K_1 K_2}{\tau_c J p^3 + J p^2 + K_1 K_2 a_2 p + K_1 K_2 a_1} \qquad (8)$$

where:

J = load inertia (slug - ft^2)

τ_c = clutch time constant (1/sec)

K_1 = amplifier-clutch coil gain $\left(\frac{\text{amps}}{\text{volts}}\right)$

K_2 = magnetic clutch gain (ft lbs/amp)

a_1 = position feedback gain = (volts/radian)

a_2 = tachometer feedback gain (volts/sec.$\times$ radian)

This transfer function can be resolved into three roots--two complex ones and one real root. The two complex roots are described by the transient specification (ζ) and (ω_n). The real root is defined as (γ). The equivalent transfer function using (γ), (ζ), (ω_n) and the required steady-state value of (e_o) (which was defined as $e_{o_{S.S.}}$), is given in equation (9).

$$\frac{e_o(p)}{e_i(p)} = \frac{e_{o_{S.S.}} \times \gamma \omega_n^2}{p^3 + [\gamma + 2\zeta\omega_n]p^2 + [2\zeta\gamma\omega_n + \omega_n^2]p + \gamma\omega_n^2} \qquad (9)$$

The synthesis procedure can be performed in two ways. First, the four performance parameters can be specified and the values of (a_1), (a_2), (K_1K_2), and (τ_c) can be computed, or a specific clutch with fixed values of (τ_c) and (K_2) can be used to compute the rise time (t_s), (a_1), (a_2), and (K_1) for specified values of (ζ) and (ω_n). If a specific clutch is used the compensation of (τ_c) is completed as the last step using equation (2) and solving for (K_o), using the criteria that [$10\tau_a < \tau_c$]. The exact value of (τ_a) is immaterial provided [$\tau_a < \tau_c/10$]. The synthesis procedure is derived by applying the Root-Locus gain and phase criteria to equations (8) and (9) with the restriction that the complex roots of equation (8) be specified by the performance parameters (ζ) and (ω_n).

The relationship between the rise time (t_s) and the real root (γ) is one of the keys to the synthesis of the servo. This relationship is given in equation (10).

$$\gamma = \zeta\omega_n + \frac{\omega_n\sqrt{1-\zeta^2}}{\text{TAN}\left(\omega_n\sqrt{1-\zeta^2}\cdot t_s + \text{TAN}^{-1}\left[\frac{\sqrt{1-\zeta^2}}{\zeta}\right] - \pi\right)} \qquad (10)$$

The real root (γ) is restricted to be greater than the real part of the complex roots which then provides a minimum and maximum obtainable value of (t_s). The minimum value of (t_s) occurs when ($\gamma \to \infty$) and corresponds to a simple second-order system. The maximum value of (t_s) corresponds to ($\gamma \to \zeta\omega_n$). The original performance specifications will provide the maximum acceptable (t_s) which in turn will specify the value of the real root (γ). There is a fixed relationship between the value of the closed-loop real root (γ) and the magnetic clutch time constant (τ_c). This relationship is given in equation (11).

$$\tau_c = \frac{1}{\gamma + 2\zeta\omega_n} \qquad (11)$$

Equations (10) and (11) provide a means for computing the required (τ_c) given a specified rise time (t_s). The restrictions on (γ) place an upper and lower limit on (t_s). The minimum and maximum values of (t_s) are given in equation (12) in terms of (ζ) and (ω_n).

$$t_{s_{min}} = \frac{\pi - \text{TAN}^{-1}\left[\frac{\sqrt{1-\zeta^2}}{\zeta}\right]}{\omega_n\sqrt{1-\zeta^2}} \qquad (12)$$

$$t_{s_{max}} = \frac{\frac{3\pi}{2} - \text{TAN}^{-1}\left[\frac{\sqrt{1-\zeta^2}}{\zeta}\right]}{\omega_n\sqrt{1-\zeta^2}}$$

An alternate synthesis procedure is to start with a given clutch in which (τ_c) is given. The value of (γ) can then be computed from equation (11), and the closed-loop servo rise time can be computed from equation (13). The real root (γ) is again constrained to be greater than ($\zeta\omega_n$).

$$\gamma = \frac{1}{\tau_c} - 2\zeta\omega_n \qquad (13)$$

$$t_s = \frac{\pi - \text{TAN}^{-1}\left[\frac{\omega_n\sqrt{1-\zeta^2}}{\gamma - \zeta\omega_n}\right] - \text{TAN}^{-1}\left[\frac{\sqrt{1-\zeta^2}}{\zeta}\right]}{\omega_n\sqrt{1-\zeta^2}}$$

With either method of fixing the value of (τ_c) the synthesis can be completed by computing a_1, a_2 and (K_1K_2). An examination of equations (10) and (11) immediately fixes the value of (a_1). This is shown in equation (14).

$$a_1 = \frac{1}{e_{o_{S.S.}}} \qquad (14)$$

The application of the Root-Locus criteria leads to the evaluation of the tachometer feedback gain (a_2). The value for (a_2) is given in equation (15) in terms of (τ_c), (ζ), ($e_{o\,s.s.}$), and (ω_n). The minimum value of (φ_1), used in equation (15),

$$a_2 = \frac{\text{TAN}\varphi_1}{e_{o\,ss}\left(\omega_n\sqrt{1-\zeta^2} + \zeta\omega_n \text{TAN}\varphi_1\right)} \qquad (15)$$

where:

$$\varphi_1 = \pi - 2\cos^{-1}(\zeta) - \text{TAN}^{-1}\left[\frac{\tau_c\omega_n\sqrt{1-\zeta^2}}{1-\tau_c\zeta\omega_n}\right]$$

is determined by setting ($\tau_c = 0$). This corresponds to the second-order system with a minimum (t_s). With given or computed values of (τ_c), (ζ), (J), (a_1), (a_2), and (ω_n), the value of (K_1K_2) can be evaluated from the Root-Locus gain criteria. The value of the product (K_1K_2) is given in equation (16).

(16)

$$[K_1K_2] = \frac{J\omega_n^2\left[(1-\tau_c\zeta\omega_n)^2 + \tau_c^2\omega_n^2(1-\zeta^2)\right]^{\frac{1}{2}}}{\left[(a_1-a_2\zeta\omega_n)^2 + a_2^2\omega_n^2(1-\zeta^2)\right]^{\frac{1}{2}}}$$

If equation (16) is used to provide clutch specifications, then (τ_a) will be chosen and the value of (K_1) can be computed by making suitable engineering choices for (L) and (K_o). If the choice for the clutch has been made, then (K_2) is known and the required value of (K_1) is computed from equation (16). (K_1) is the closed-loop gain of the driving amplifier and the value of (K_o) must be derived from the restrictions on (τ_a), R and h for the fixed value of (K_1) and (τ_c). Equation (17) shows this relationship.

$$K_o = \frac{K_1 L}{\tau_a} \Rightarrow \text{CLUTCH GIVEN} \qquad (17)$$

$$K_1 = \frac{\tau_a K_o}{L} \Rightarrow \text{TO SPECIFY CLUTCH}$$

where:

R, L, τ_c, L, K_2, AND $[K_1K_2]$ KNOWN;
CONSTRAINT: $\tau_a < \tau_c/10$

If specifications for a clutch are to be made, then (τ_c) is computed from (t_s) and the value of (K_1) is determined from equation (17) by assuming nominal values of (K_o) and (L).

The design of the excitation coil amplifier with a nominal gain (K_o) involves a locus of values for (R) and (h). The acceptable values of (R) and (h) are determined by the value of (K_1) and (τ_a).

Amplitude Characteristics With Respect to Rise Time (t_s)

The normalized amplitude of the transient response at the rise time (t_s) is a function of the closed-loop parameters (ζ), (γ) and (t_s). For near optimum values of (ζ) the amplitude is near the first crossing of its steady-state value. Values of (ζ) less than (.5) would probably not be chosen for normal servo synthesis, but the equations developed in the synthesis section are equally useful for analysis provided the exact relationship between the transient amplitude at the rise time (t_s) for the prescribed range of (ζ) is known. The rise time (t_s) has a minimum and maximum value for each pair of (ζ) and (ω_n). By computing the transient amplitude as a function of a fixed percentage increase of (t_s) over its minimum value, or a fixed percentage decrease over its maximum value, the amplitude can be expressed independently of (ω_n).

Figure 5 shows the percent difference between the amplitude of the transient and the steady-state value as a function of ($t_s/t_{s\,min}$), ($t_s/t_{s\,max}$), and (ζ). Figure 6 shows the value of (γ/ω_n) as a function of (t_s) and (ζ). Figure 7 shows the relationship between ($t_{s\,min}/\omega_n$) and ($t_{s\,max}/\omega_n$) for various values of (ζ). An examination of Figure 5 shows that the percent difference in response amplitude is at least an order of magnitude smaller than the accuracy of the hardware components that would constitute the position servo.

Servo Stability Considerations [4,5]

In actual practice, the amplitude-phase characteristics of dry-powder magnetic clutches differ from a first-order lag because of magnetic hysteresis.

Magnetic hysteresis appears as a fixed phase lag at very low frequencies and approximates a constant time delay at high frequencies. The transfer function of the clutch at high frequencies has a phase lag which is porportional to the frequency. For these reasons the only complete approximation of the magnetic clutch is a meromorphic function. However, the amplitude characteristics at servo signal frequencies approximates a first-order lag very closely and the magnitude of the transfer function approaches -6 to -8 db per octave at high

frequencies. The main difference between the actual clutch characteristics and the equivalent first-order lag is the phase characteristic. The clutch model used in this synthesis procedure cannot produce an unstable servo. Since the physical clutch has a monotonic increasing phase lag, the position servo can be unstable. Figure 8 shows a typical amplitude-phase plot of a dry-powder magnetic clutch. Indicated in Figure 8 is the first-order lag (τ_c) which would be used in the synthesis procedure. The first-order lag (τ_c) is constructed from matching the amplitude of clutch transfer function to a theoretical lag without considering the phase plot. The equivalent first-order lag (τ_c) should have a slightly higher or equal amplitude at signal frequencies than the clutch. The phase lag of the actual clutch is less than the equivalent first-order lag (τ_c) for all frequencies less than (ω_a). At ($\omega = \omega_a$) the phase lag of the clutch and the equivalent lag (τ_c) are equal. For typical clutches (ω_a) is greater than or equal to ($3/\tau_c$). These relationships are shown in Figure 8.

The synthesis procedure developed in this paper will always provide a stable closed-loop servo with actual clutches, provided the value of (ω_n) is constrained to be less than or equal to (ω_a), and (t_s) is kept smaller than ($.9t_{s\,max}$). These relationships are placed in their proper perspective when it is realized that (t_s) would be taken as near ($t_{s\,min}$) as possible, and (ω_n) will usually be less than ($1/\tau_c = \omega_c$), the equivalent corner frequency of the clutch.

References

1. Magnetic Servo Clutch, February, 1951, Boeing Airplane Company, H. Lusissar and E. N. Zangen

2. Characteristics of Some Magnetic-Fluid Clutch Servomechanisms, 1949, MIT Servomechanisms Laboratory, A. J. Parziale and P. D. Tilton

3. Investigation of Some Design Fundamentals of the Magnetic Particle Clutch, July, 1950, MIT Servomechanism Laboratory, J. P. Ivaska, R. Kramer, and G. C. Newton, Jr.

4. The Theory and Behavior of the Lear Magnetic Dry-Powder Clutch Type Servo, September, 1957, Lear Incorporated, A. Bepristis

5. Basic Feedback Control System Design, 1958, McGraw-Hill, C. J. Savant, Jr.

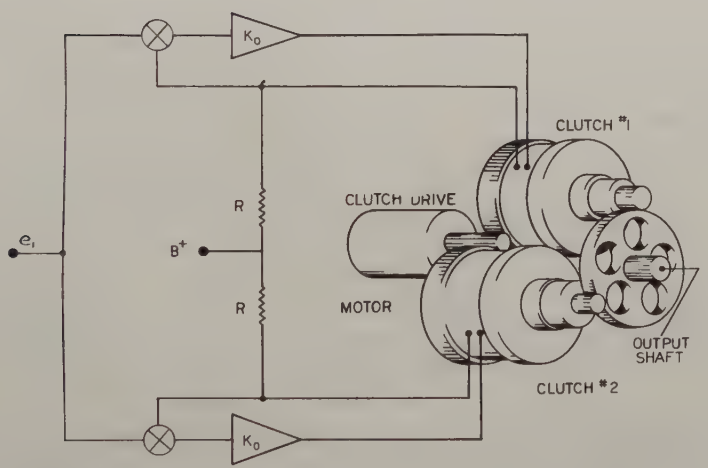


Fig. 1. Electrical-mechanical configuration of amplifier and clutch unit.

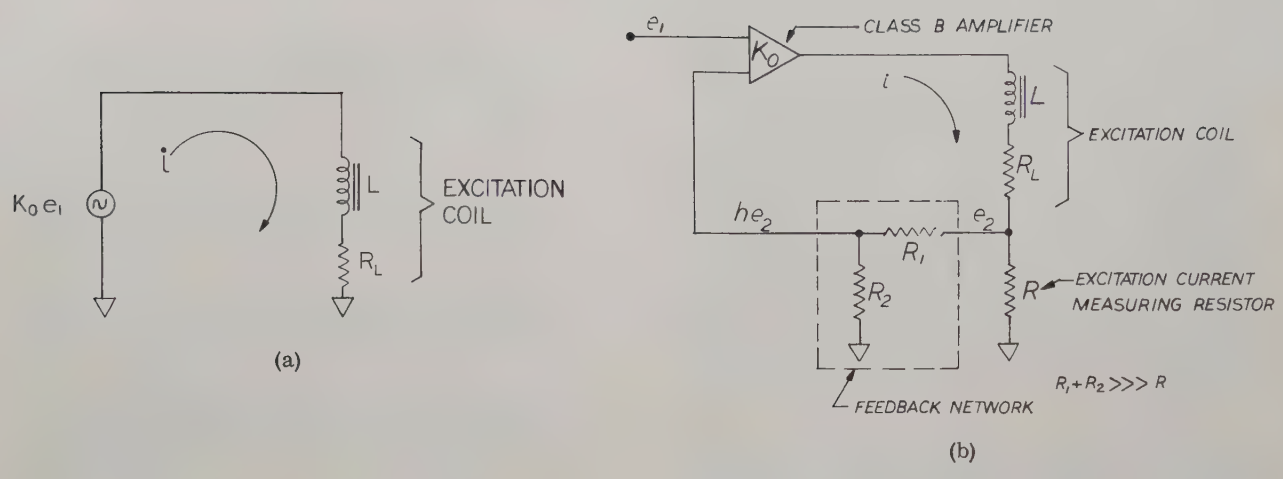


Fig. 2. (a) Excitation coil equivalent electrical circuit.
(b) Equivalent circuit of Fig. 1 including coil time lag compensation.

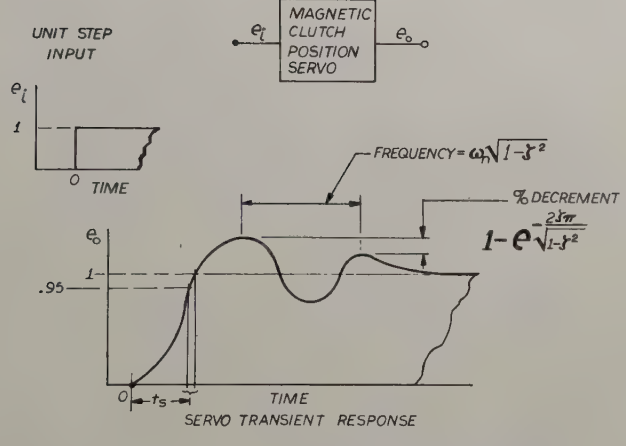


Fig. 3. Closed loop performance specifications.

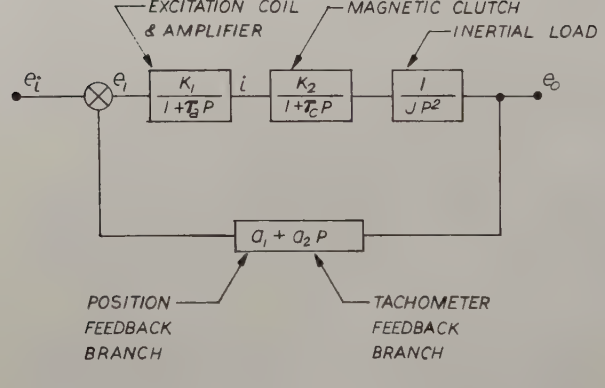


Fig. 4. Complete closed loop servo with magnetic clutch actuator.

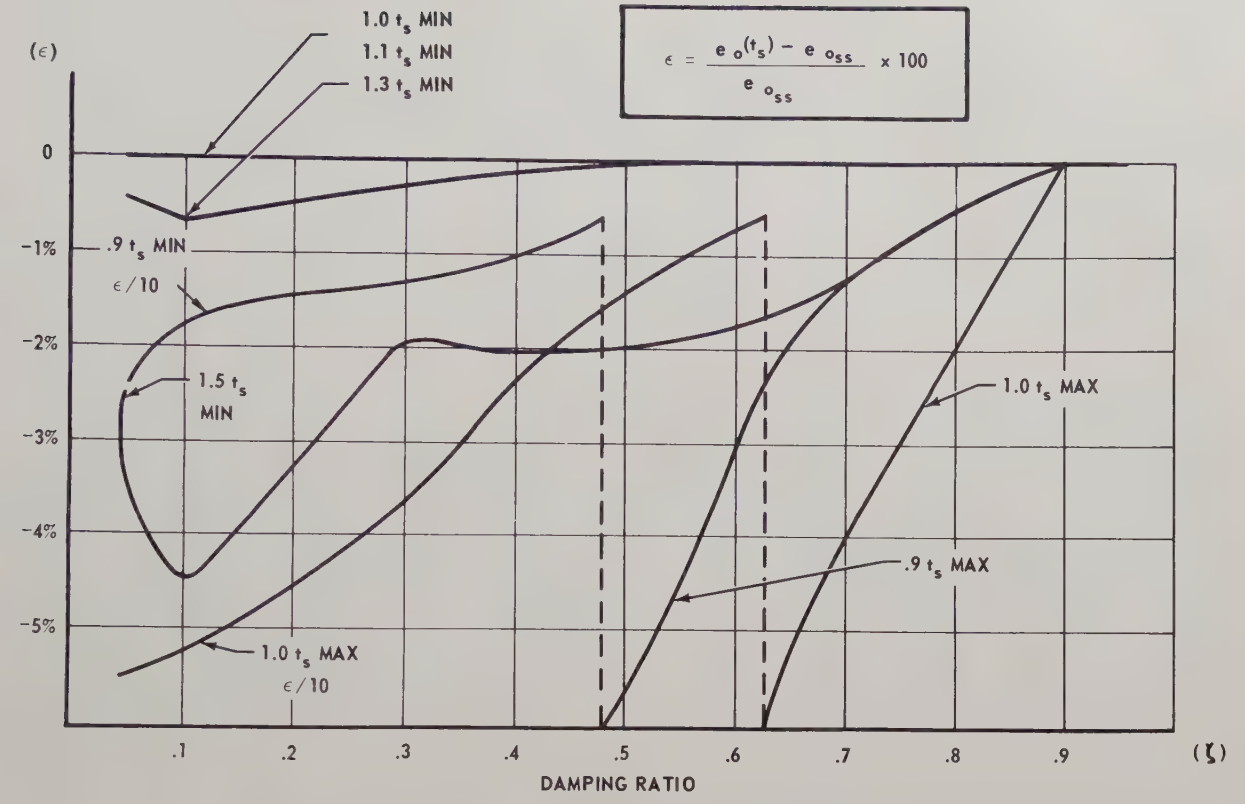


Fig. 5. Transient response at rise time (t_S) with respect to the steady-state value $Co_{S.S.}$.

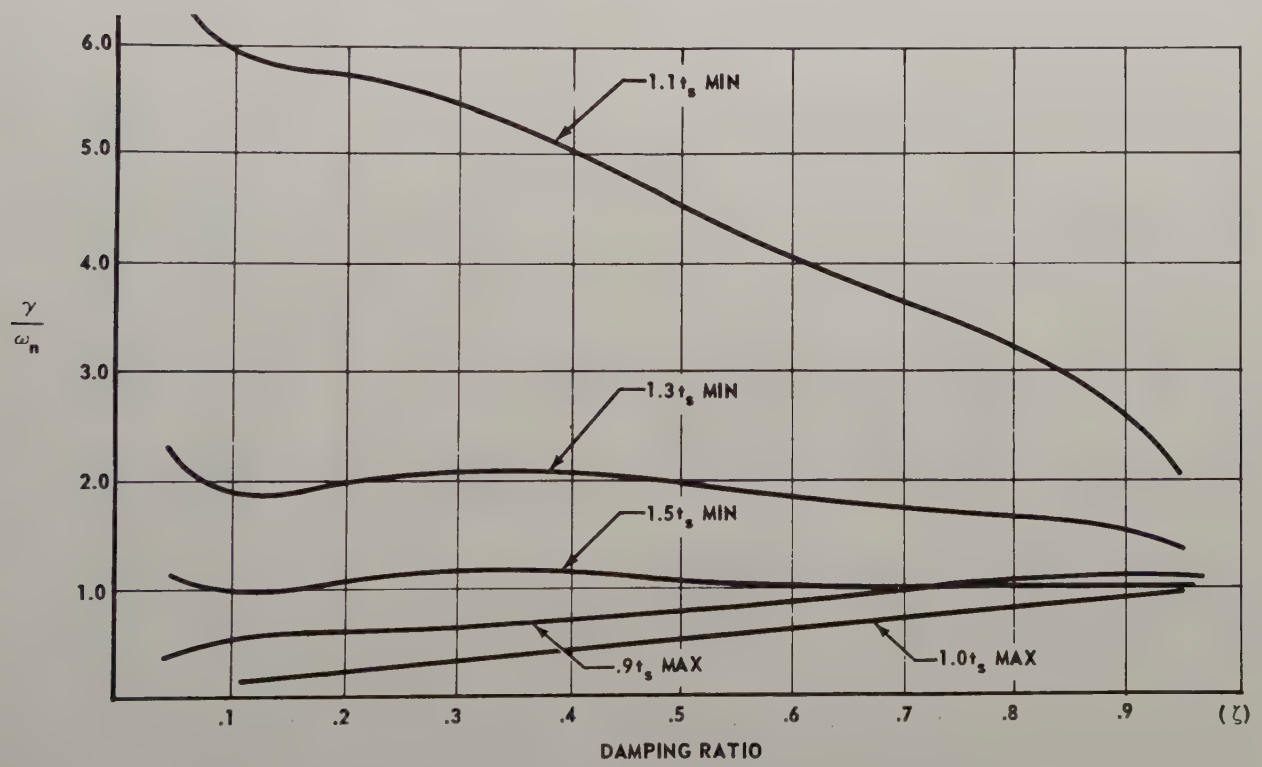


Fig. 6. Normalized real root (γ) with respect to rise time (t_S) and damping ratio (ζ).

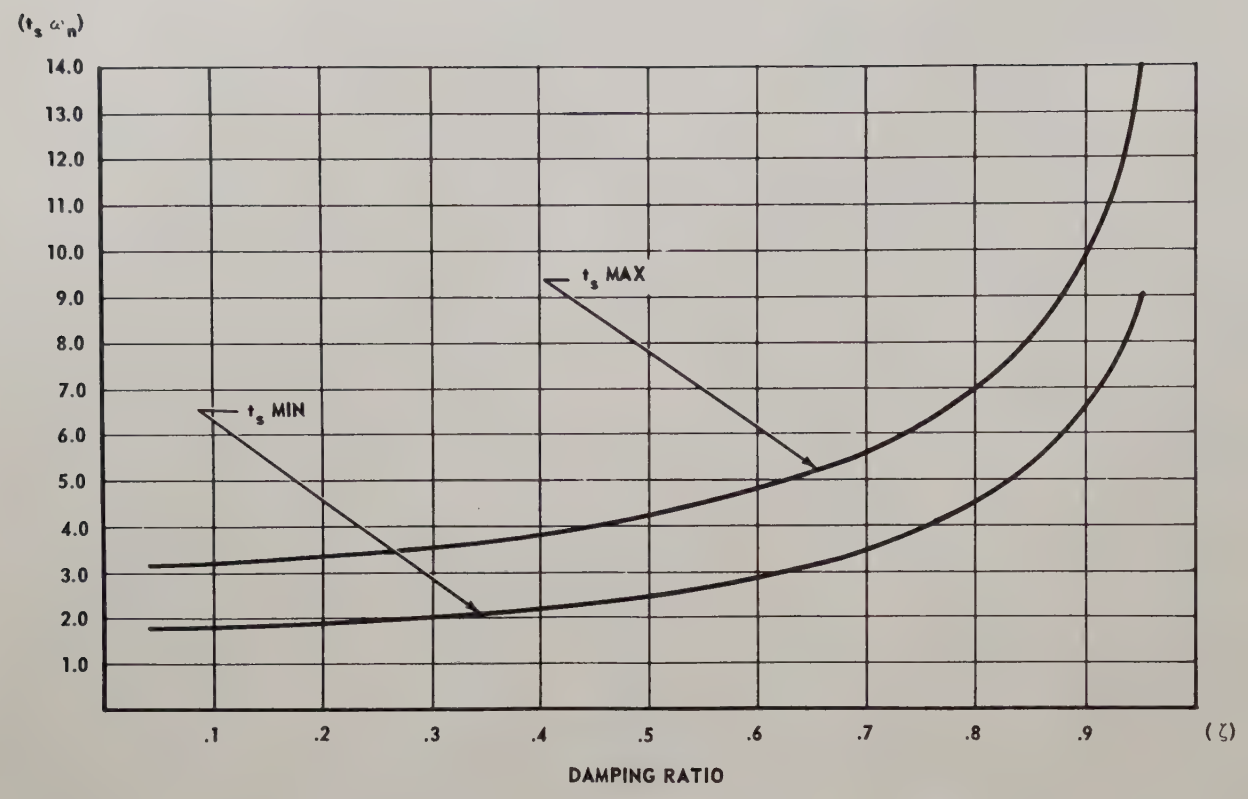


Fig. 7. Normalized minimum and maximum rise time vs damping ratio.

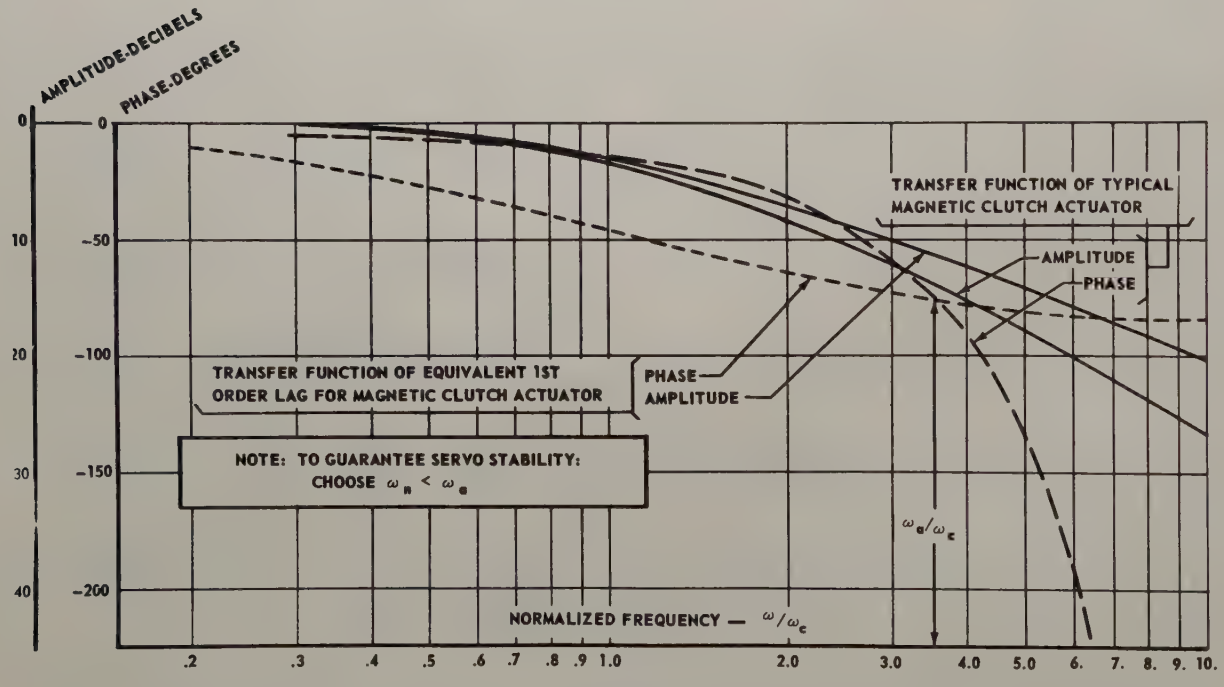


Fig. 8. Typical magnetic clutch actuator transfer function and equivalent theoretical first-order model.

SYNTHESIS OF A SELF ADAPTIVE AUTOPILOT FOR A LARGE ELASTIC BOOSTER

by
George W. Smith
Design Engineer, Advanced Electronics
Avionics Department
The Martin Company
Denver 1, Colorado

Abstract

The design of a conventional autopilot requires a detailed knowledge of the structural dynamics of the missile that is not easily obtainable. The self-adaptive autopilot considered in this paper would not depend on this detailed knowledge. This system would, ideally, adjust itself to maintain optimum performance in an environment of changing inertial, aerodynamic, and structural parameters. The mechanization of a suitable self adaptive system is complicated by the difficulty of getting a continuous measure of system performance. The method used, in this study, to obtain the basic modes is based on decomposition of the transient response.

Introduction

In an autopilot system for a large flexible booster with relatively low bending frequencies and small structural damping the system parameters, such as rigid body natural frequency and damping, should be invariant to changes in the mass and aerodynamic environment of the vehicle. The design of a conventional autopilot involves a detailed knowledge of the structural dynamics of the missile that is not easy to obtain. Therefore an autopilot system must be evolved that does not depend on a detailed knowledge of the missile structural characteristics.

The concept of the self adaptive autopilot has been applied in an attempt to solve these problems. Ideally, this system automatically adjusts itself to maintain optimum performance in an environment of changing inertial, aerodynamic, and structural parameters. The mechanization of a suitable self adaptive system is complicated by difficulties associated with the continuous measure of system performance. A method based on decomposition of the transient response to obtain the basic modes is described.

Satisfactory self adaption can be accomplished if information on performance can be obtained with negligible time delay. This requirement necessitated the development of a method for obtaining system performance parameters instantaneously. This method is based on the transient response of the system to periodic impulse disturbances. In general this response will consist of the rigid body modes plus the lightly damped elastic modes. The concept of the self tuning frequency sensor has been evolved to accomplish the decomposition. This device is excited by the missile's impulse response and continuously furnishes information about the rigid body natural frequency and damping. The frequencies of the elastic modes are also continuously furnished. Using this information, the rigid body dynamics of the missile are self adapted to be invariant with changing inertia and aerodynamic environment. The frequency information on the elastic modes is used to adjust a compensating filter that keeps the structural feedback in the autopilot to a negligible level.

This system can satisfactorily eliminate the structural feedback problem in a missile system including two elastic bending modes. The ability of the system to handle two bending modes having been demonstrated, the system can theoretically be extended to handle any number of bending modes.

The Elastic Missile System

Consider the block diagram of the general elastic missile shown in Figure 1, where

$$G_E = \prod_{k=1}^{n} \frac{\gamma_k s^2 + \beta_k s + 1}{a_k s^2 + b_k s + 1} .$$

The closed loop transfer function for this system is

$$\frac{\theta_o}{\theta_i} = \frac{AG_E}{s^2 + AG_E (K_P + K_R s)} .$$

Substituting for G_E,

$$\frac{\theta_o}{\theta_i} = \frac{A \prod_{k=1}^{n} \frac{\gamma_k s^2 + \beta_k s + 1}{a_k s^2 + b_k s + 1}}{s^2 + A (K_D + K_R s) \prod_{k=1}^{n} \frac{\gamma_k s^2 + \beta_k s + 1}{a_k s^2 + b_k s + 1}} .$$

Letting n = 2 and plotting the amplitude ratio and phase of this equation for typical values of the damping and natural frequencies associated with the elastic missile, it becomes evident that any theory for self-adaption which takes care of the elastic modes must be capable of distinguish-

ing between these modes (Fig. 2). For example, if a self adapted filter were successfully set up to cancel out the first bending mode characteristic, the problem would only have been transferred to the second bending mode, and so forth. Thus, any solution to the self adaption of the elastic missile must be capable of taking care of several modes simultaneously.

The self adaptive system must be able to:

1) Distinguish between the various elastic modes;

2) Adjust a filter for cancellation of each individual mode.

To keep the problem within bounds, it was decided to study a rigid missile plus two bending modes. This is sufficient to demonstrate a method applicable for the general case where the system must:

1) Distinguish between the first and second bending modes;

2) Adjust a suitable filter for cancellation of these modes.

Philosophy of Overall System

The general approach to the solution of the problem is as follows:

The rigid body mode is made self adaptive. The purpose of this feature is to establish the rigid body dynamics equal to some predetermined invariant factor. This provides the missile with uniform dynamics over the flight trajectory. The missile impulse response is then passed through a rejection filter that rejects the rigid body mode. This leaves a residue that consists of a superposition of the elastic mode responses. Since these modes possess essentially zero damping, the elastic residue can be expressed as

$$\theta_{oe} = \sum_{k=1}^{n} A_k \sin(\lambda_k t + \psi_k).$$

The elastic residue is decomposed to obtain the component frequencies λ_k. The purpose of this operation is to obtain signals for the positioning of cancellation filters. Thus, since the elastic dynamics enter into the problem according to

$$G_E = \prod_{k=1}^{n} \frac{\gamma_k s^2 + \beta_k s + 1}{a_k s^2 + b_k s + 1},$$

a cancellation filter is reasonably employed that is essentially the reciprocal of this quantity. Examining the individual contributions from G_E (Fig. 2) shows that each mode has the effect of contributing a tall spike at the frequency λ_k. Thus the approach to cancellation is based on measuring the frequency λ_k and positioning a rejection filter at the frequency λ_k. This can be done having only a knowledge of λ_k.

In the determination of the elastic mode frequencies, the concept of the self tuning frequency sensor has been evolved. This device can be started with arbitrarily set values. It can tune itself by information received from the elastic residue to give out signals proportional to the component frequencies λ_k in the elastic residue. If the problem requires knowledge of n bending frequencies, n parallel channels of the device are employed. The theory presented is for two channels, but the device can easily be extended to more.

The operation of the proposed autopilot system for two elastic bending modes is illustrated by the block diagram in Figure 3. As illustrated in this figure, the impulse response of the missile is applied to the self tuning frequency sensor. Since the rigid body mode is the lowest frequency, the first channel of this device selects and measures the rigid body natural frequency. A separate measurement of damping ratio is made, the theory of which is developed later. The values obtained for these quantities are then compared with the desired values and error signals obtained. These error signals are used to close a loop around the autopilot position and rate gains K_D and K_R to maintain the rigid body natural frequency and damping ratio invariant. After rejection of the self adapted rigid body mode the remaining output signal represents the impulse response resulting from the first and second bending modes, the elastic residue. This elastic residue is the input to the second channel of the bending mode frequency sensor. The output from this channel is the square of the frequency of the first bending mode rejection filter, positioning this filter to suppress the first bending mode dynamics. Simultaneously the first channel output positions a rejection filter for cancellation of the first bending mode from the input to the second channel of the frequency sensor. This channel works in exactly the same manner as the first channel, and selects the second bending mode frequency.

The Frequency Sensor

The differential equation describing harmonic motion is

$$\ddot{\theta} + \lambda^2 \theta = 0.$$

The solution of this differential equation is a sine wave

$$\theta = A \sin \lambda t.$$

Thus, a sine wave and the second derivative of this wave are related by the square of the frequency. This principle can be used as a basis for measuring the frequency of an unknown sine wave.

In the case of the elastic missile, the transient response after removing the rigid body motions is the sum of sinusoids of unknown frequency and amplitude. A device is desired that is able to sort out these sinusoids and give information about the frequencies contained in the sum. The problem is solved under the additional restriction of allowing no prior knowledge of the component frequencies.

Suppose the sum of sinusoids is run through a device whose output would be predominantly a sinusoid of the lowest frequency contained in the sum, no matter what this frequency may be. The output of this device is then differentiated twice and the result of the double differentiation compared with the original. The magnitude of the two signals would be in the ratio of the frequency squared. Suppose further that the signal λ_1^2 is used to tune the original input filter to further favor the lowest frequency. In the limit, the frequency of the lowest frequency sinusoid in the sum can be measured. Having a measure of λ_1^2, this measure can be used to tune a rejection filter so, if the original signal is run through this filter, the lowest frequency component will be knocked out. If this signal is fed to a second channel similar to the first one described, it will select the sinusoid of next highest frequency, λ_2, to favor and, in the limit, will measure λ_2^2. This procedure can be carried on as far as required.

For details of the device, first consider the input filter that favors the lowest harmonic present. If this filter is made a simple undamped second order system, it has the transfer function

$$\frac{\theta_o}{\theta_i} = \frac{k}{1 + \tau s^2}$$

where τ is adjustable by λ_1^2. The amplitude characteristic for this filter is the second order characteristic with zero damping (Fig. 4).

Suppose initially $1/\tau$ is set to some value lower than any expected frequency present in the sum of sinusoids. Then the portion of the curve considered is on the right half of the infinite spike. This portion of the curve is asymptotic to -12 db/octave. This will provide a two to one separation of amplitudes if the first and second frequencies are of equal magnitude and are separated by two to one. Since, in the missile, the magnitude of the first bending will exceed that of the second, the output of the filter will favor the lower frequency.

Assuming that the frequency can be measured, the filter can be tuned to further favor the lower frequency. If this process is carried to the limit for the system shown, infinite response to the lowest mode will result. Since this is undesirable, automatic gain control is used to limit the amplitude of this oscillation to some predetermined level. This is done by feeding the output of the filter into a deadspace device, the width of which is equal to the double amplitude of the desired oscillation. If the oscillation exceeds the width of the deadspace, pulses will come out of the deadspace that can be fed back to limit the amplitude of the oscillation out of the filter. This introduces some distortion of the sine wave, but experience has not shown this to be serious.

The measurement of the frequency of the amplitude limited oscillation will now be studied. Consider the sine wave $\theta = A \sin \lambda t$. Differentiating this function twice $\ddot{\theta} = A\lambda^2 \sin \lambda t$. Then $\frac{\ddot{\theta}}{\theta} = -\lambda^2$. To accomplish the double differentiation, a network is used with a transfer function of

$$\frac{\theta_o}{\theta_i} = \frac{s^2}{(1 + \tau_1 s)^2}.$$

The lag term in the denominator is used to avoid having infinite gain at high frequencies, which would result in instability of the circuit. To provide the proper phase relations during the comparison process, the direct signal is passed through a network whose transfer function is

$$\frac{1}{(1 + \tau_1 s)^2}.$$

If the direct signal were now to be multiplied by λ^2 and summed with the double differentiated signal the result would be zero. Suppose, however, that it is multiplied by some quantity, λ'^2, that in general is not equal to λ^2. The sum of the direct signal and the double differentiated signal is then a sine wave whose magnitude is equal to the error $\epsilon = \lambda'^2 - \lambda^2$. If this is integrated with respect to time the result would again be zero. Therefore, the two signals are rectified before they are compared. This then gives the absolute value of the two waves, and on comparison a string of pulses is found. The pulses are positive or negative as ϵ is positive or negative. If these pulses are integrated and the output of the integrator called λ^2 the stable condition for this integrator is $\lambda'^2 = \lambda^2$, or $\epsilon = 0$. This gives the quantity for adjustment of the input filter and to multiply the direct signal by. A block diagram of one channel of the frequency sensor is shown in Figure 5.

The Self Adapted Autopilot

Consider now an autopilot system with provi-

sions for sensing and cancelling out two elastic bending modes. Such a system with no compensation for the elastic properties would be represented by the block diagram of Figure 2 and n = 2 in the term G_E. Examining Figure 2, the open loop frequency response curves for one environment for this system corresponding to a typical large elastic booster, this system is seen to have a negative phase margin when the transfer function crosses the 0 db axis, and hence is unstable. The normal approach to this problem is to gather data allowing examination of the stability of this system under all anticipated environmental conditions of the missile, and then design a compensating network to achieve stability. This procedure is objectionable from several points of view:

1) It requires detailed knowledge of the anticipated missile structural and environmental characteristics. This is at best a very dubious and laborious task.

2) It requires a compromise in the compensating network that attempts to stabilize the loop over the expected structural and environmental ranges. This at best provides marginal performance. The compensating network that will make the loop stable over the trajectory the missile is going to fly is the result of a calculated guess. It is impossible to optimize the performance under these conditions.

These objectionable problems can now be eliminated. The first can be eliminated if the elastic modes contained in the missile response are sensed and this information used to adjust compensation to provide stable operation of the loop for any and all environments. If the elastic modes are effectively eliminated and nullified, only the rigid body response is left. If the rigid body portion of the autopilot is now made self adaptive, optimum performance of the missile response can be provided that is invariant to environmental conditions.

As pointed out previously the elastic modes enter into the loop through a transfer function of the form

$$G_E = \prod_{k=1}^{n} \frac{\gamma_k s^2 + \beta_k s + 1}{a_k s^2 + b_k s + 1}$$

The constants in this product are so related that each term contributes a tall spike in the open loop transfer function, the net contribution in phase being zero. Figure 6 shows a typical term.

The altitude of the peak is of the order of +28db, due to the structural damping ratio of 0.02 or less. If the location of the spike were known compensation could be used to provide cancellation, that is essentially the reciprocal of the mode transfer functions. Thus the approach is to sense the frequency of the bending modes and position a rejection filter in the loop to cancel the elastic structural feedback. Such a filter is shown in Figure 7. The success of the whole scheme thus depends on sensing the elastic frequencies λ_k.

A block diagram of a self adapted autopilot for a large elastic booster, using the self tunin frequency sensor previously described, is shown i Figure 3. This figure shows a system for sensing and eliminating two elastic bending modes.

So far the self adaption of the rigid body mode has not been described. This is done by manipulating the rigid body rate and displacement K_R and K_D respectively. In the class of vehicle being considered, the response characteristics of th inner loops (such as the hydraulic loop) and the dynamic characteristics of the sensing instrument: is very high as compared to the inertial response characteristics of the missile. Thus, these response characteristics are separated by a factor o ten to one. Under these conditions, the dynamics of the inner loop and sensing instruments can be ignored and the assumption made that these element are essentially ideal. This leaves only inertia and aerodynamic forces as far as the rigid body is concerned. Under these assumptions the rigid tran fer function becomes

$$\frac{\theta_o}{\theta_i} = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}.$$

If the elastic effects have been sensed and nullified, this becomes the overall closed loop transfer function of the missile autopilot system. In a space environment, aerodynamic effects become negligible and the damping and restoring forces on the missile are determined by the rate and displacement gains. To self adapt this transfer function, the frequency ω_n is sensed in a manner similar to that described for the frequency sensor.This measured frequency is then compared with some desired set value for this frequency and an error ϵ_{ω_n} obtained. This error is fed into an integrator whose output is K_D. Thus, the integrator either increases or decreases K_D until ω_n equals its desired set value. A simple procedure can be followed for the adjustment of the rigid body damping ratio. Consider the differential equation

$$\ddot{\theta}_o + 2\xi\omega_n\dot{\theta}_o + \omega_n^2\theta_o = \omega_n^2\theta_i.$$

Now suppose ξ_o is the value of the damping ratio desired, and let $\xi = \xi_o + \Delta\xi$. Then

$$\ddot{\theta}_o + 2(\xi_o + \Delta\xi)\omega_n\dot{\theta}_o + \omega_n^2\theta_o = \omega_n^2\theta_i.$$

Solving for $\omega_n\dot{\theta}_o\Delta\xi$, we have,

$$\omega_n\dot{\theta}_o\Delta\xi = \omega_n^2(\theta_i - \theta_o) - 2\xi_o\omega_n\dot{\theta}_o - \ddot{\theta}_o.$$

Since ω_n is fixed at its known value, by manipulating K_D, and hence is known; theoretically $\Delta\xi$

could be calculated, if θ_o, θ_i, $\dot{\theta}_o$, and $\ddot{\theta}_o$ were known, by performing a division. This seems objectionable, because it involves measuring four quantities and performing a division. What is desired is to measure two quantities and not perform a division. This can be done as indicated in Figure 8.

The missile output θ_o is measured and $\dot{\theta}_o$ and $\ddot{\theta}_o$ are generated by differentiation. A lag term is introduced in the differentiations to cut off the high frequency and provide stable operation. The output of the summing junction is a quantity, $\frac{\omega_n \dot{\theta}_o \Delta\xi}{(1+\tau s)^2}$ that can vary in sign, not only with $\Delta\xi$ but also with $\dot{\theta}_o$. Hence, this quantity is multiplied by the output from the first differentiator to give a term whose sign varies only as $\Delta\xi$. The multiplier provides a quantity whose magnitude and sense is a measure of $\Delta\xi$. If this quantity is integrated and the output of the integrator made K_R, then this integrator either increases or decreases K_R until

$$\Delta\xi = 0.$$

Thus, the autopilot system is self adapted to have dynamic response characteristics that are invariant to the missile environmental conditions.

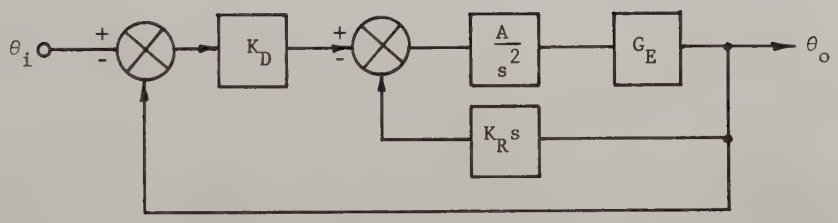


Fig. 1. Elastic missile autopilot block diagram.

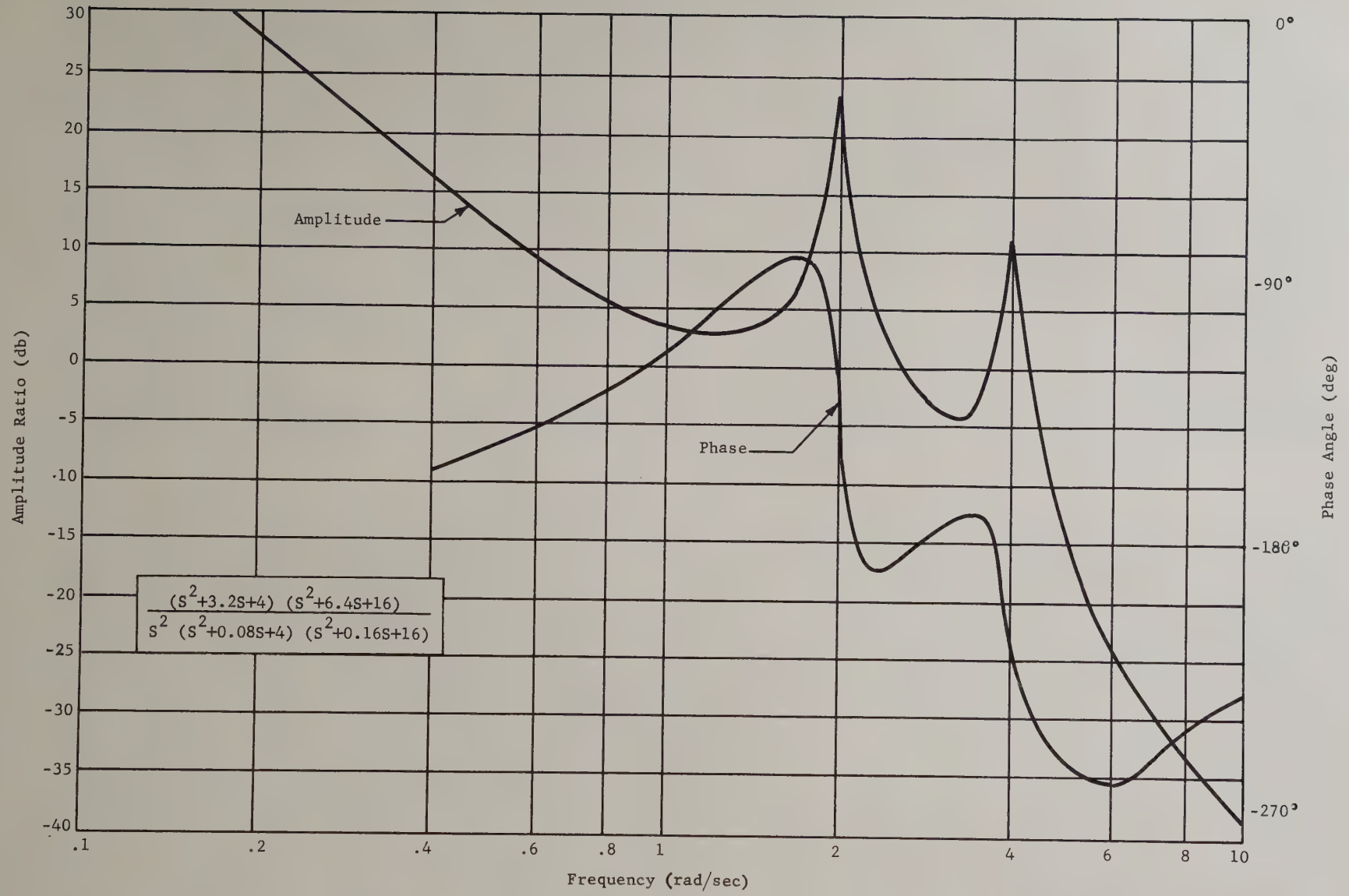


Fig. 2. Typical open loop frequency response for elastic missile.

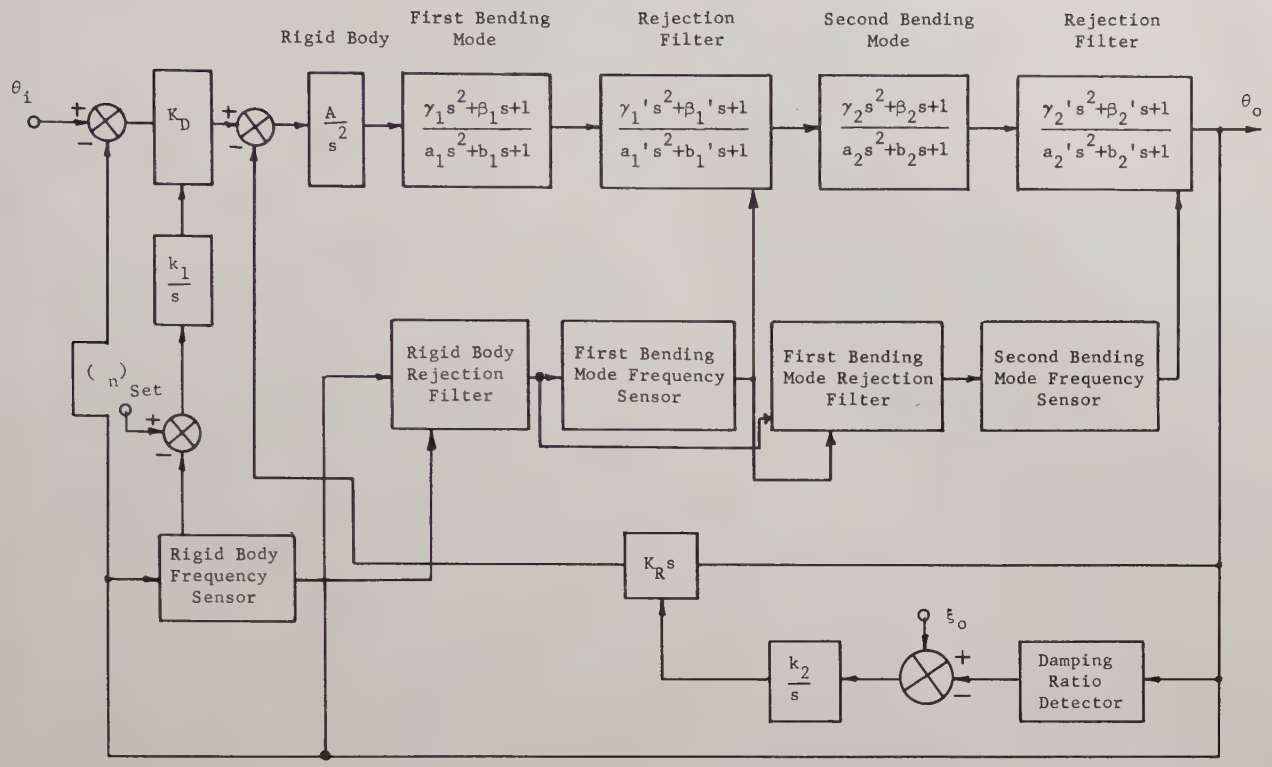


Fig. 3. Block diagram of self adapted autopilot for elastic missile.

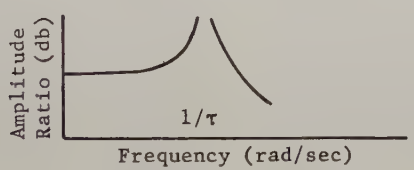


Fig. 4. Input filter characteristic.

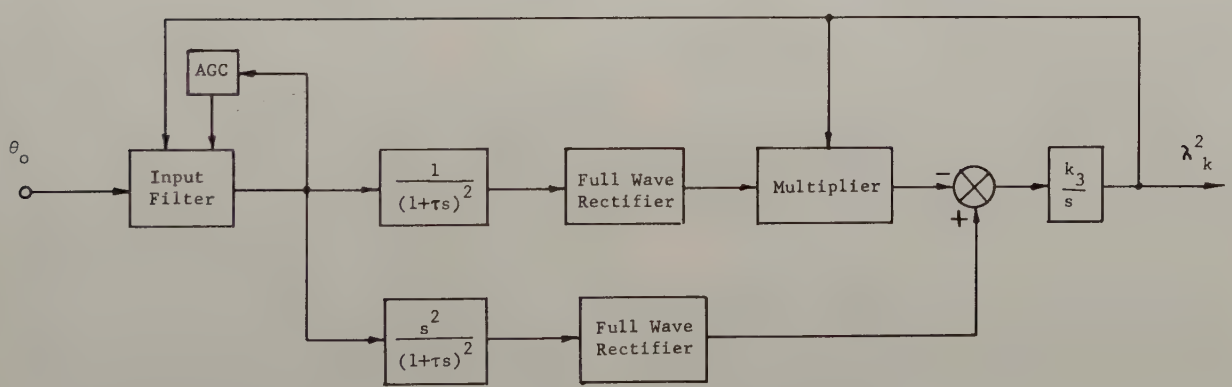


Fig. 5. Block diagram of self tuning frequency sensor.

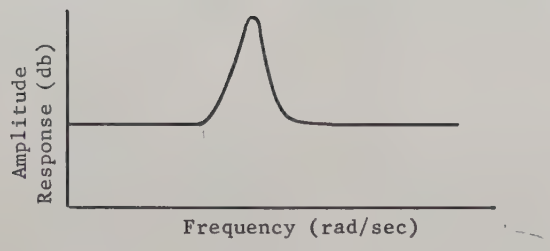


Fig. 6. Amplitude ratio for one elastic mode.

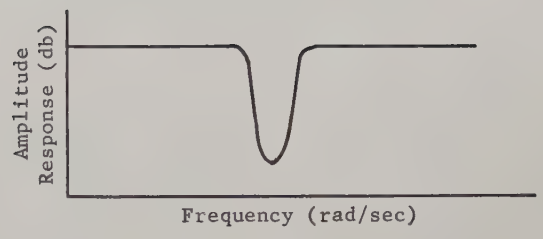


Fig. 7. Rejection filter characteristic.

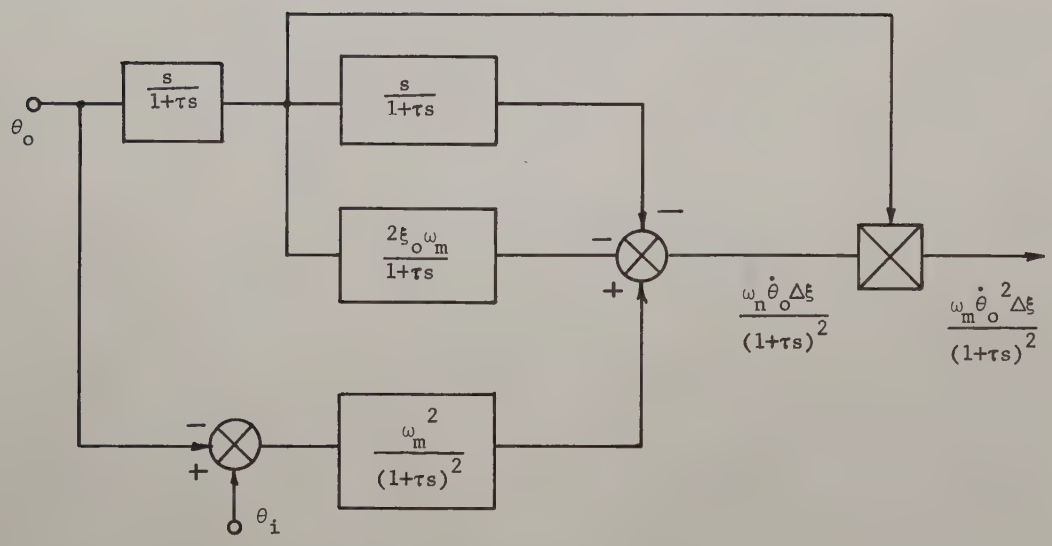


Fig. 8. Damping ratio error detector.

DESIGN OF OPTIMUM BEAM FLEXURAL DAMPING IN A MISSILE BY APPLICATION OF ROOT-LOCUS TECHNIQUES

Ronald J. Hruby
Northrop Corporation, Norair Division
Hawthorne, California

Summary

The control of a missile in free-fall in the presence of external disturbances requires that the control system be stable when various body bending modes are excited. It is an inevitable consequence of the internal and external missile environment that all body bending modes are excited. The general method of stabilizing a missile is to choose a minimum of inertial sensors such as rate and attitude gyros, and through the use of passive electrical networks in the information paths, generate stabilizing poles and zeros in the control system open-loop transfer functions.

Generally, the electrical shaping networks are approximations to "maximally" flat functions with the cutoff frequency below the first bending mode resonant frequency, and in extreme cases, a notch filter is needed to eliminate body bending information. The basic purpose of the electrical shaping networks is to attenuate the high signal frequencies circulating in the control loop without the loss of important low frequency data. In some applications the use of passive shaping networks becomes very difficult because of the large attenuation required near the first body bending resonant frequency and the required amplitude flatness at lower frequencies. If passive shaping cannot be used, then active networks must be substituted. This is an undesirable complexity.

This paper presents an optimum design procedure to utilize the characteristic properties of completed airframe and missile structures to acquire a maximum increase in modal damping without auxiliary equipment and further autopilot complexity by the proper choice of the sub-assembly attaching fixtures.

Completed airframes and missiles are composed of a very lightweight integral structure in which a large number of equipment sub-assemblies are attached. The sub-assemblies or equipment modules are generally massive and of small volume so that they act as a rigid mass attached to the lightweight integral structure via an equivalent lateral spring constant and viscous damper. For example, the guidance package and the rocket engine of a ballistic missile are usually attached fore and aft of the missile integral tank structure. Each of these elements can be quite rigid and will act as a separate mass from the integral tank structure. The proper choice of the equivalent lateral spring constant and viscous damper of the attaching fixtures that hold the sub-assemblies to the basic lightweight missile or airframe integral structure will result in an optimum increase in modal damping for a particular beam-flexural mode of the completed structure.

To determine the values of the proper equivalent spring constant and viscous coefficient, the dynamic principles underlying the Frahm damper are used by extending the analysis of the Frahm damper concept to include a general linear flexural beam which satisfies a linear partial differential equation which can be solved by the product of the solutions of two ordinary, linear differential equations.

The extension of these principles is accomplished by deriving the solution of the attached sub-assembly to the integral structure through normal servomechanism analysis. This analysis uses the analytical relationships of the Evans or Root-Locus method.

Introduction

The design of air-space and outer-space vehicles usually results in light, rigid structures which have the dynamic properties of beams with a high finesse ratio. The combination of rigidity, minimum weight, and slenderness means the first body bending mode, and possibly some of the other modes will have a very low damping ratio which will be on the order of 10% or less.

The controls engineer is faced with the problem of attaching inertial elements to the vehicle to provide flight attitude stability. Typical propulsion units such as rocket and jet engines that are used to drive air-space or outer-space vehicles also excite the structural flexural modes.

The controls engineer must design his closed-loop control system so that the vibrational energy in the body-bending modes is not increased as the result of closing the attitude control loop. In effect, the control system must contain signal attenuation at body-bending frequencies so that the open-loop gain is less than one. The required amount of signal attenuation at body-bending frequencies increases very rapidly as the flexural mode damping ratio approaches zero. These requirements have led to overly complex autopilot systems with less than the desired reliability.

This paper presents a passive method by which the damping ratio for any beam flexural mode can be optimally increased by taking advantage of the sub-system equipment modules which are attached to the missile body. In a missile, the rocket engine assembly, the heavy payload package, or various electronic and guidance packages can be used to improve the damping ratio of any of the body-flexural modes by choosing the attaching structural characteristics by the procedure outlined in this paper. The analysis applies only to lateral bending modes since the longitudinal modes are relatively unimportant.

The procedure for improving modal damping is based upon an extension of the Frahm damper principle to that of a free-fall beam with concentrated masses attached to the beam with an equivalent lateral spring and dashpot.

Since the concentrated mass is necessary for the system operation and the spring is required for structural reasons, the dashpot is the only new element added to the missile system. The size of the equivalent dashpot is controlled by the concentrated mass and the modal resonant frequency. The weight penalty of the dashpot appears to be a negligible minimum.

Description of a Frahm Damper System

There are many mechanical structures which appear to be a large mass located at the end of a firmly imbedded spring. Such a structure is shown in Figure 1.

If a forcing function is applied to the suspended mass with a forcing frequency near the spring-mass resonant frequency, the resulting position amplitudes become extremely large and therefore undesirable. Theoretically, a forcing frequency equal to the spring-mass resonant frequency results in infinite position amplitudes. In a practical case, the mass is constrained by stops. The forcing function is usually an external disturbance on which no control can be exerted. The problem is to find a method of linearly restricting the positional amplitude of the mass without adding a damping device in parallel with the spring. The solution of this problem is the addition of a small mass, coupled with a spring and a dashpot to the large mass, with a natural frequency very close but not equal to the natural frequency of the large mass and spring. Such a system is shown in Figure 1 and is called a Frahm damper.[1]

The analysis of the simple system in Figure 1 leads to equation (1), which gives the damping ratio (ζ_M) of the system mass (M) in terms of the system parameters (K), (B), (κ), and (m).

$$\zeta_M \simeq \frac{1}{2}\left[\frac{m}{2M+m}\right]^{\frac{1}{2}} \qquad (1)$$

$$B = mM\left[\frac{3K}{2(m+M)^3}\right]^{\frac{1}{2}}$$

$$\kappa = \frac{mK}{M}\left[\frac{M}{m+M}\right]^{\frac{1}{2}}$$

where:

ζ_M = Damping ratio of mass (M)

K = Spring constant of mass (M) (lbs/ft)

m = Frahm mass (slugs)

κ = Frahm spring constant (lbs/ft)

B = Frahm dashpot coefficient (lbs. sec/ft)

The Frahm mass (m) is arbitrary. In general, the larger the Frahm mass (m) the greater the increase in the effective damping of the large mass (M), but the Frahm mass (m) is usually restricted by other known considerations to be much smaller than (M).

Application to a Free-Beam

The usefulness of a Frahm damper to increase the modal damping of a free-beam and therefore a missile in free-fall, arises from the flexural characteristics of a beam which makes the differential equations of motion similar to the system shown in Figure 1. Figure 2 shows how a Frahm damper is applied to a free-free beam. The damper is applied at a point (X_1) on the beam and is "tuned" to the mode of the beam which requires the increase in modal damping. The application of a Frahm damper to a free-free beam is like a servo feedback system in which the feedback gain is proportional to the normal mode amplitude squared and the mass of the Frahm damper. Since weight is precious for a missile, it requires the damper to be located at the point of maximum modal amplitude, usually at the tail or the nose (either end of the beam). The analysis for a beam does not result in a set of simple equations such as exhibited in equation 1. However, a set of simultaneous equations do result, which when solved give the same information required for designing the damper. The analysis of the problem is based upon a non-homogenous beam with low modal damping. The solution is derived from knowledge of the beam's modal frequency (ω_n), inherent damping (B_1), orthogonal mode shape ($\psi_n(x)$), modal mass (M_n), and the mass of the Frahm damper (m). The analysis is based upon the application of the Root-Locus theorems.

Mathematical Analysis

The basic differential equation of motion of a free-free beam is given by equation 2. This equation applies to the system shown in Figure 2. The solution of this equation depends upon the capability of forming a product solution of time and lateral displacement.

The solution in time is taken to be a sinusoid at the driving frequency of the forcing function (i.e., the forcing function is taken to be the product of a sinusoid in time with a variable magnitude in (X)). This beam loading is illustrated in Figure 3.

Equation 2 applies to a non-homogeneous beam, bending in one plane with no gyroscopic or cross-dimensional coupling effects but including inherent beam damping.

$$\frac{\partial^2}{\partial x^2}\left[E(x)\cdot I(x)\cdot\frac{\partial^2}{\partial x^2}y(x,t)\right] + \rho(x)B_1\frac{\partial}{\partial x}y(x,t) + \rho(x)\frac{\partial^2}{\partial t^2}y(x,t) = \rho(x)f(x,t) \quad (2)$$

where:

$F(x)$ = Young's modulus (lbs/ft)

$I(x)$ = Cross-sectional moment of inertia $(ft)^4 = R^2 \cdot S$

R = Cross-sectional radius of gyration $(ft)^2$

S = Cross-sectional area $(ft)^2$

B_1 = Damping force/mass/unit length (lbs.sec.ft./slug ft.)

$\rho(x)$ = mass/unit length (slugs/ft)

$f(x,t)$ = External force/mass/unit length (lbs.ft/slug)

The lateral displacement y(x,t) is a product of a function of time and a function of (x). The forcing function is also the product of two independent functions. Equation (3) shows this relationship.

$$y(x,t) = y_1(x)\cdot y_2(t) \quad (3)$$

$$f(x,t) = f_1(x)\cdot f_2(t)$$

The function $y_1(x)$, in general, is a sum of independent functions of (x) each multiplied by an arbitrary constant which is evaluated by incorporating the boundary conditions of the beam and the specific forcing function $f_1(x)$. Equation (4) gives the general form $y_1(x)$.

$$y(x) = \sum_{n=1}^{\infty} g_n \psi_n(x) \quad (4)$$

where:

g_n = A modal constant which is derived from $f_1(x)$, $\psi_n(x)$ and M_n(modal mass)

$\psi_n(x)$ = Orthogonal N-mode shape

$\psi_n(x)$ is evaluated either by the analytical solution of the beam differential equation when possible or by experimentally recording the mode shape which is the most generally used method. The analytical expression for $\psi_n(x)$ has as many unknowns as the independent boundary conditions of a free-free beam. These boundary conditions are that the third and fourth derivatives of the lateral displacement at both ends of the beam are zero. This gives four unknown constants in the analytical expression for $\psi_n(x)$. These constants are evaluated by applying the four boundary conditions and the orthogonality properties of the individual modes. A normalizing constant called the modal mass (M_n) is defined in terms of the mode shape and the linear beam mass density $\rho(x)$. Equation (5) shows the relationship between (M_n), (g_n), $\psi_n(x)$ and the driving function $f_1(x)$.

$$M_n = \int_0^{\ell} \rho(x)\cdot\psi_n^2(x)\cdot dx \quad (5)$$

$$g_n = \frac{1}{M_n}\int_0^{\ell}[\rho(x)\,\psi_n(x)\,f_1(x)]\,dx$$

The exact dimensional values of (M_n) and $\psi(x_i)$ are unimportant because only the ratio $[\psi(x_i)\cdot\psi(x_j)/M_n]$ is used in control system transfer functions. When the mode shape is determined experimentally, the value of Mn is computed in the non-dimensional coordinates defining $\psi_n(x)$. This makes the use of the computed (M_n) mandatory with the particular modal coordinates used.

Substituting equations (3), (4) and (5) into equation (2) by invoking the orthogonality properties of the differential equation and taking the Laplace transform of equation (2) with respect to time, gives the solution for $y_2(p)$ which is the Laplace transform of $y_2(t)$. The result of these analytical steps is given in equation (6) which is the transformed total solution.

$$Y(x,p) = \sum_{n=1}^{\infty} \frac{g_n\,\psi_n(x)\,F_2(p)}{p^2 + B_1 p + \omega_n^2} \quad (6)$$

where: $F_2(p)$= Laplace transform of $f_2(t)$

$$\omega_n^2 = \frac{1}{M_n}\int_0^{\ell}\frac{\partial^2}{\partial x^2}\left[E(x)\,I(x)\frac{\partial^2\psi_n(x)}{\partial x^2}\right]\psi_n(x)\,dx$$

For a non-uniform bean the integral defining (ω_n^2) is a constant and (ω_n^2) is the scaling factor for orthogonality.

The general solution as given in equation (6) for a free-free beam with an experimentally derived mode shape is the basis for the application of the Frahm damper. The forcing function

f(x,t) is derived by solving the differential equation of motion of a Frahm damper with a position input and an inertial reaction force on the beam as the output. This allows the computation of gn and $F_2(p)$ for a particular mode. Due to the resonant low pass characteristics of a damped spring mass system, the improvement in body modal damping can only occur for one mode per Frahm damper assembly. Figure 2 shows the system including the input ($\ddot{Z}(t)$) and output ($f_R(t)$) functions of the Frahm damper. The transfer function relating the Laplace transforms of $f_R(t)$ to $\ddot{Z}(t)$ is used in an equivalent feedback system to derive the solution for the n^{th} mode. The solution provides a means of adjusting the components of the Frahm damper to optimally increase the damping ratio of the n^{th} mode. The transfer function for $f_R(t)$ driven by $\ddot{Z}(t)$ is given in equation (7), which also shows the differential equations of motion of the Frahm damper of Figure 2.

$$\text{(a)}\quad f_R(t) = (Z_1 - Z)k + B(\dot{Z}_1 - \dot{Z}) = m\ddot{Z}_1 \qquad (7)$$

$$m\ddot{Z}_1 = (Z - Z_1)k + B(\dot{Z} - \dot{Z}_1) \quad \text{Differential Equations}$$

$$\text{(b)}\quad \frac{F_R(p)}{Z(p)} = -\frac{p^2(k + Bp)}{p^2 + \frac{B}{m}p + \frac{k}{m}} \quad \text{Transfer function for the reaction force}$$

where: $F_R(p)$ = Laplace transform of $f_R(t)$

$Z(p)$ = Laplace transform of $Z(t)$

The equivalent feedback system of the Frahm damper and the missile body is shown in Figure 4 and only applies to the n^{th} mode. The final transfer function relating the lateral beam displacement to the forcing function for a beam with an attached Frahm damper at point (X_1) is given in equation (8).

$$Y_n(x,p) \quad \frac{g_n F_2(p)\Psi_n(x)[p^2 + (B/m)p + k/m]}{[p^2 + B_1 p + \omega_n^2][p^2 + \frac{B}{m}p + \frac{k}{m}] + [\frac{m\Psi_n^2(x_1)}{Mn}]p^2[\frac{k}{m} + \frac{B}{m}p]} \qquad (8)$$

Application of Root-Locus Theorems

From Figure (4) the equivalent open-loop transfer function GH(p) can be derived and is given in equation (9).

$$GH(p) = \frac{m\Psi_n^2(x_1)}{Mn} \frac{p^2[(B/m)p + k/m]}{[p^2 + B_1 p + \omega_n^2][p^2 + (B/m)p + k/m]} \qquad (9)$$

The Root-Locus criteria state that GH(p) must equal (-1) at each of the closed-loop roots. The variables $\Psi_n(x_1)$, (M_n), and (B_1) are known because the dynamic properties of the missile are known and (m) is the mass of the Frahm damper and is known because it is the mass of one of the sub-system equipment packages. The values of (k) and (B) must be chosen so that the damping of the closed-loop roots have a maximum increase in magnitude over (B_1) for a given magnitude of (m).

The Root-Locus theorems are applied to equation 9 along with additional constraints that guarantee that the closed-loop roots have a maximum increase in damping over the unmodified beam equivalent of the missile body. These additional constraints are applied to the shape of the Root-Locus as (m) increases from zero and to the location of the open-loop pole and zeros which depend upon (k/m) and (B/m). Figure 5 shows the Root-Locus plot for equation 9. The locus of the root leaving the original beam root (P_3) is constrained to be an arc of a circle. The maximum decrease in (ϕ) is required for a minimum change in (ℓ_3) to maximize the increase in root damping ratio for a fixed (m). This means $(d\Delta\phi/d\ell_3)$ must be a maximum value. Equation 10 shows this evaluation

$$\frac{d\Delta\phi}{d\ell_3} = \frac{1}{R\cos\frac{\Delta\phi}{2}}\left[1 - \frac{\ell_3}{R}\frac{dR}{d\ell_3}\right] \qquad (10)$$

where:

$\cos(\phi_1 - \Delta\phi) = \zeta_0$ for closed-loop roots $(R_1 \& R_2)$

$\frac{dR}{d\ell_3} = 0$ for maximum (ζ_0) for a given $(\Delta\phi)$

The derivative $(d\Delta\phi/d\ell_3)$ has a maximum value if $(dR/d\ell_3) = 0$. Therefore, the locus of the root (R_1) must be an arc of a circle starting at (P_3) with a radius (ω_n) as shown in Figure 5. To facilitate the analysis, the parameters (k/m) and (B/m) are normalized with respect to (ω_n). The zero and pole locations, and the zero and pole distances, from (R_1) are given by equation (11) for Figure 5 and for equation (10).

$$Z_1 = \frac{\omega_0^2}{\tau_0}\omega_n;\quad Z_2 = 0;\quad Z_3 = 0 \qquad (11)$$

$$P_1 = \left[-\frac{\tau_0}{2} + i\sqrt{\frac{4\omega_0^2 - \tau_0^2}{4}}\right]\omega_n$$

$$P_2 = \left[-\frac{\tau_0}{2} - i\sqrt{\frac{4\omega_0^2 - \tau_0^2}{4}}\right]\omega_n$$

$$P_3 = \left[-\frac{\tau_1}{2} + i\sqrt{\frac{4 - \tau_1^2}{4}}\right]\omega_n$$

$$P_4 = \left[-\frac{\tau_1}{2} - i\sqrt{\frac{4 - \tau_1^2}{4}}\right]\omega_n$$

$$\left[\frac{L_1}{\omega_n}\right]^2 = [\sin(\phi - \Delta\phi)]^2 + \left[\frac{\omega_0^2}{\tau_0} - \cos(\phi - \Delta\phi)\right]^2$$

$$\left[\frac{\ell_1}{\omega_n}\right]^2=\left[\left[\omega_o^2-\left(\frac{\tau_o}{2}\right)^2\right]^{\frac{1}{2}}-\sin(\phi-\Delta\phi)\right]^2+\left[\frac{\tau_o}{2}-\cos(\phi-\Delta\phi)\right]^2$$

$$\left[\frac{\ell_4}{\omega_n}\right]^2=\left[\cos(\phi-\Delta\phi)-\cos\phi\right]^2+\left[\sin\phi+\sin(\phi-\Delta\phi)\right]^2$$

$$\left[\frac{\ell_3}{\omega_n}\right]=2\sin\left[\frac{\Delta\phi}{2}\right]$$

$$\left[\frac{\ell_2}{\omega_n}\right]^2=\left[\left[\omega_o^2-\left(\frac{\tau_o}{2}\right)^2\right]^{\frac{1}{2}}+\sin(\phi-\Delta\phi)\right]^2+\left[\frac{\tau_o}{2}-\cos(\phi-\Delta\phi)\right]^2$$

where: $\omega_o=\frac{1}{\omega_n}[k/m]^{\frac{1}{2}}$; $\tau_o=\frac{1}{\omega_n}[B/m]$; $\tau_1=\frac{1}{\omega_n}[B_1]$

The eight angles shown in Figure 5 are related to the variables describing the pole-zero lengths of equation (11). Equation (12) shows these relationships.

(12)

$$\epsilon_1=\frac{\pi}{2}-\phi \qquad \epsilon_3=\frac{\pi}{2}-\frac{\Delta\phi}{2}$$

$$\text{Tan}\,\epsilon_1=\frac{\sin(\phi-\Delta\phi)-\left[\omega_o^2-\left(\tau_o/2\right)^2\right]^{\frac{1}{2}}}{\tau_o/2-\cos(\phi-\Delta\phi)}$$

$$\text{Tan}\,\epsilon_2=\frac{\tau_o/2-\cos(\phi-\Delta\phi)}{\left[\omega_o^2-\left(\tau_o/2\right)^2\right]^{\frac{1}{2}}+\sin(\phi-\Delta\phi)}$$

$$\text{Tan}\,\epsilon_4=\frac{\cos(\phi-\Delta\phi)-\cos\phi}{\sin\phi+\sin(\phi-\Delta\phi)}$$

$$\text{Tan}\,\theta_1=\frac{\sin(\phi-\Delta\phi)}{\omega_o^2/\tau_o-\cos(\phi-\Delta\phi)}$$

$$\cos\phi=\frac{\tau_1}{2}=\frac{B_1}{2\omega_n}=\zeta_1 \qquad \sin\frac{\Delta\phi}{2}=\frac{\ell_3}{2\omega_n}$$

$$\cos(\phi-\Delta\phi)=\zeta_o \quad \text{(Damping of closed-loop root } (R_1))$$

The pole-zero lengths and the angles of Figure 5 are constrained by the gain and phase criteria which guarantee that (R_1) is a root of equation (8). The gain and phase criteria are applied to the variables of equations (11), and (12) in equation (13).

(a) $$\theta_1=\epsilon_4+\epsilon_3+\epsilon_1-\epsilon_2-2\left[\frac{\pi}{2}-\phi+\Delta\phi\right] \quad (13)$$

(b) $$G_o=\frac{\ell_1\ell_2\ell_3\ell_4}{\tau_o L_1}$$

where: (a) Root-Locus phase criterion

(b) Root-Locus gain criterion

$$G_o=\frac{m\,\psi_n^2(x_1)}{Mn}$$

The solution for (τ_o), (ω_o), and (ζ_o) as a function of (G_o), (ω_n), and (ζ_1) is determined by substituting the pole-zero lengths of equation (11) and the angles of equation (12) into the gain and phase criteria of equation (13).

The solution is not obtainable in closed form because the equations for (τ_o) and (ω_o) are not linear, nor do they contain common polynomial factors which might allow simplification.

The equations are solvable by a simple concise iterative cycle. A further optimization on the solution for (τ_o) and (ω_o) must take place during the iterative cycle. The location of (P_1) must be approximately the same distance from the closed-loop root (R_1) as the original beam root (P_3). This means that (ℓ_3) $\simeq$ (ℓ_1). Attaching the Frahm damper to the beam introduces an additional pair of complex roots (R_3) and (R_4). By constraining the value of (ℓ_1) to be slightly larger than (ℓ_3), and the range of (ϵ_1) to be less than ($\pi/2$), the complex roots (R_3) and (R_4) will have a slightly larger damping ratio than the closed-loop roots (R_1) and (R_2). These iterative constraints are exhibited in equation (14).

$$\ell_1 \geq \ell_3 \ (\ell_1 \text{ approaches } \ell_3) \qquad (14)$$

$$\epsilon_1 \leq \frac{\pi}{2} \ \begin{matrix}\text{(defines the solvable)}\\ \text{(range of the problem)}\end{matrix}$$

If (ϵ_1) $>\frac{\pi}{2}$ or (ℓ_1) $<$ (ℓ_3) then (R_3) and (R_4) are possibly less damped than (R_1) and (R_2). This constraint defines the range in (ϕ) and ($\Delta\phi$) in which the optimizing constraints hold. The minimum value of ($\phi-\Delta\phi$) is approximately 65° which covers the expected values of (ζ_1) and (m).

Iteration Cycle for (K) and (B)

The iteration cycle begins by choosing a value for ($\Delta\phi$) which automatically fixes (ℓ_3), (ϵ_1), and (ϵ_3) since (ϕ) and (ω_n) are given. A value for (ϵ_1) is chosen arbitrarily. The value of (ϵ_4) is computed from (ϕ) and ($\Delta\phi$). Since (ℓ_1) is approximately equal to ℓ_3 , both (ω_o) and (τ_o) can be determined from the equations for (Tan ϵ_1) and (ℓ_1), which then provides the means for computing (ϵ_2). The value of (θ_1) is computed from the phase criteria equation (13a). The value of (ω_o^2/τ_o)$_a$ is computed from the known parameters (θ_1), sin ($\phi-\Delta\phi$) and cos ($\phi-\Delta\phi$). Both (ω_o) and (τ_o) are known from assuming the value of (ϵ_1). The value of (ω_o^2/τ_o)$_b$ is computed from (τ_o) and (ω_o). A comparison is made between (ω_o^2/τ_o)$_a$ and (ω_o^2/τ_o)$_b$. If (ω_o^2/τ_o)$_a$ is greater than (ω_o^2/τ_o)$_b$ then a second value of (ϵ_1) is chosen which is larger than the first value. An estimate can be made on the increment in (ϵ_1) by using (ω_o^2/τ_o)$_b$ to compute a value of (θ_{1b}). Since (θ_{1b}) is larger than (θ_1) the difference between (θ_{1b}) and (θ_1) is added to the original

value of (ϵ_1) to start the second iteration cycle. The iterative steps per cycle are outlined below.

Iteration Cycle

1. Given (ϕ), fix $(\Delta\phi)$ and choose a value of (ϵ_1) to begin the iteration or take $(\epsilon)_i$ from the i^{th} cycle to commence the $(i+1)$ cycle.

$$\epsilon_4 = \tan^{-1}\left[\frac{\cos(\phi-\Delta\phi)-\cos\phi}{\sin(\phi)+\sin(\phi-\Delta\phi)}\right]$$

$$\frac{\ell_3}{\omega_n} = 2\sin\left(\frac{\Delta\phi}{2}\right)$$

2. Set the ratio $\tan(\epsilon_1)_i = \left[\frac{\alpha}{\beta}\right]$ on a slide rule and choose specific values of (α) and (β) so that $(\alpha^2+\beta^2) \simeq (\ell_3/\omega_n)^2$

3. Compute $(\tau_o)_i$:

$$(\tau_o)_i = 2\left[\beta + \cos(\phi-\Delta\phi)\right]$$

4. Compute $(\omega_o^2)_i$

$$(\omega_o)_i^2 = \left[\alpha + \sin(\phi-\Delta\phi)\right]^2 + \frac{(\tau_o)_i^2}{4}$$

5. Compute $(\epsilon_2)_i$:

$$(\epsilon_2)_i = \tan^{-1}\left[\frac{\beta}{\alpha + 2\sin(\phi-\Delta\phi)}\right]$$

6. Compute $(\theta_1)_i$

$$(\theta_1)_i = \epsilon_4 + (\epsilon_1)_i + \phi - \frac{\pi}{2} - (\epsilon_2)_i - \frac{3}{2}\Delta\phi$$

7. Compute $(\omega_o^2/\tau_o)_{ai}$:

$$(\omega_o^2/\tau_o)_{ai} = \frac{\sin(\phi-\Delta\phi)}{\tan[\theta_1]_i} + \cos(\phi-\Delta\phi)$$

8. Compute $(\omega_o^2/\tau_o)_{bi}$:

$$(\omega_o^2/\tau_o)_{bi} = (\omega_o^2)_i/(\tau_o)_i$$

9. If $(\omega_o^2/\tau_o)_{ai} > (\omega_o^2/\tau_o)_{bi}$ then choose a new value of $(\epsilon_1)_{i+1}$ larger than $(\epsilon_1)_i$
To estimate $(\epsilon_1)_{i+1}$: Compute $(\theta_1)_{bi}$

$$(\theta_1)_{bi} = \tan^{-1}\left[\frac{\sin(\phi-\Delta\phi)}{\left(\frac{\omega_o^2}{\tau_o}\right)_{bi} - \cos(\phi-\Delta\phi)}\right]$$

and

$$(\epsilon_1)_{i+1} = (\epsilon_1)_i + (\theta_1)_{bi} - (\theta_1)_i$$

10. If $(\omega_o^2/\tau_o)_{ai} < (\omega_o^2/\tau_o)_{bi}$ choose a new (ϵ_1) smaller than $(\epsilon_1)_i$ by step 9, in which $(\theta_1)_{bi} < (\theta_1)_i$

11. If $(\omega_o^2/\tau_o)_i < 1$ then estimate $(\epsilon_1)_{i+1}$ by setting $(\epsilon_1)_{i+1} = (\epsilon)_i + \frac{(\theta_1)_{bi} - (\theta_1)_i}{2}$

12. With the new value of $(\epsilon_1)_{i+1}$ re-enter the iteration cycle at step 2.

The completed iteration produces accurate values of (τ_o) and (ω_o), for each pair of the fixed values of (ϕ) and $(\Delta\phi)$, which satisfy the equations (11), (12) and (13) as well as the optimizing constraints. The corresponding value of (m) can then be determined because (ω_n), (ℓ_1), (ℓ_2), (ℓ_3), (ℓ_4), (L_1), (τ_o), and $(\psi_n^2(x_1)/Mn)$ are known. This relationship is given in equation (15).

$$m = \frac{Mn\,\ell_1\,\ell_2\,\ell_3\,\ell_4}{\psi_n^2(x_1)\,\tau_o\,L_1} \qquad (15)$$

In practice, the value of (m) is given along with the other properties of the beam. The required inverse functional relationship was determined by using the iteration procedure for the ranges of (ϕ) and (ϵ_1) consistent with the optimization and solvability constraints and plotting the results versus (G_o). The result of this analytical procedure is given in Figures 6, 7, and 8. Figure 6 shows the Frahm damper normalized dashpot coefficient (τ_o) as a function of the equivalent open-loop gain (G_o) for various values of initial beam damping ratio (ζ_1). Figure 7 shows the Frahm damper normalized spring constant (ω_o) as a function of the equivalent open-loop gain (G_o) for the range in (ζ_1). Finally, Figure 8 gives the damping ratio (ζ_1) of the least damped, closed-loop root of the modified beam as a function of (G_o).

Conclusions

To demonstrate the usefulness of the functional relationships shown in Figures 6, 7, and 8, a Frahm damper is applied to a theoretical beam whose flexural properties are given in Table 1.

Table 1
Sample Beam Flexural Properties

M_n	Mode	ω_n	ζ_1	ψ_n(Nose)	ψ_n(Tail)
900 slugs	1	30	.009	2.0	1.5
700 slugs	2	70	.010	-1.0	1.6

Table 1 could apply to a missile whose lift-off weight lies between 75,000 lbs. and 150,000 lbs. A typical major sub-system module might weigh as much as 640 lbs. which corresponds to 20 slugs. The flexural properties of the resultant Frahm-damper-modified beam, with the damper applied to the nose, is given in Table 2.

Table 2*
Flextural Properties of Modified Sample Beam

Mode	(ζ_o)	(ω_o)	(τ_o)	(m) slugs	(B) lbs sec/ft	(K) lbs/ft
1	.1593	.918	.504	20	303	15,200
2	.0932	.968	.325	20	456	88,600

*(damper applied at nose)

A comparison of the initial and final values of the damping ratio (i.e., (ζ_1) compared with (ζ_0)) indicates a substantial improvement can be realized from the Frahm damper principle. The spring constant for the first mode as indicated in Table 2 is relatively small compared to typical lateral spring constants, but would be acceptable with its matched dashpot.

A dashpot with a viscous coefficient of 500 (lbs. sec/ft) and driven at a frequency of 40 radians/sec with an input position amplitude of 1.00 inches would provide a maximum reaction force of 1670 lbs. If the piston area is 1.0 (sq in), then the maximum internal pressure would be 1670 psi with a required piston stroke from (2) to (4) inches. A dashpot with these characteristics would weigh no more than 10 lbs. Considering the potential improvement in beam flexural damping this weight penalty would be negligible.

References

1. Mechanical Vibrations, D. Hartog
2. Control Engineers Handbook, J. G. Truxal
3. Servomechanism Practice, W. Ahrendt, C. J. Savant, Jr.
4. Principles of Inertial Navigation, C. J. Savant, Jr., C. B. Solloway, R. C. Howard, C. A. Savant

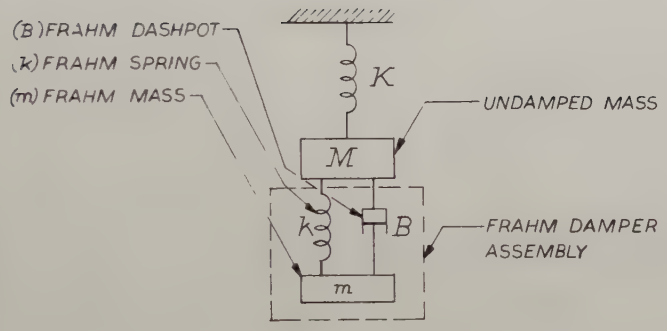


Fig. 1. Undamped mass with a Frahm damper.

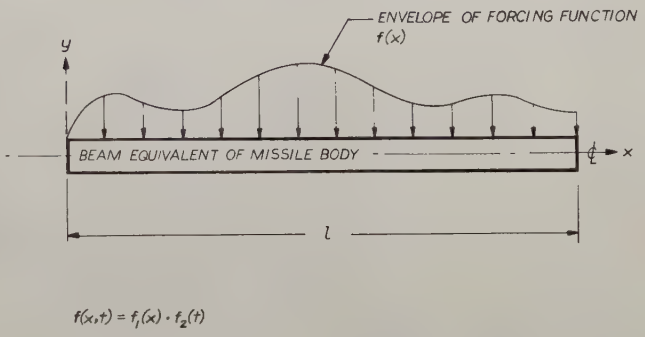


Fig. 3. Force function loading of a missile body.

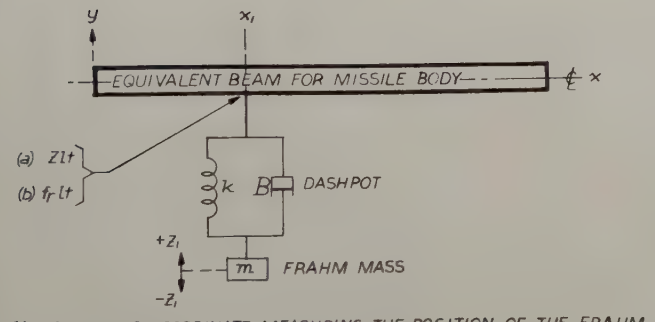


Fig. 2. Frahm damper attached to a free-free beam.

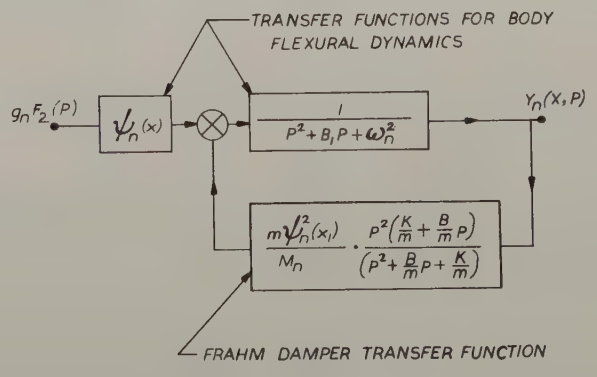


Fig. 4. Equivalent feedback system for a missile body.

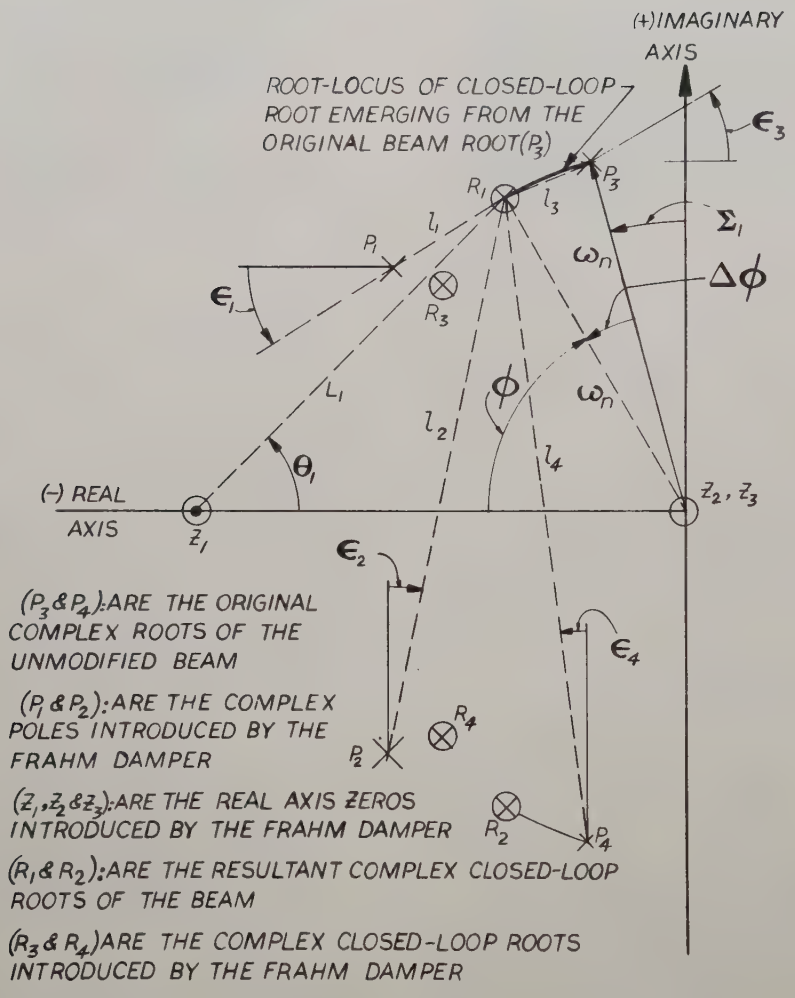


Fig. 5. Root-locus of beam with attached Frahm damper.

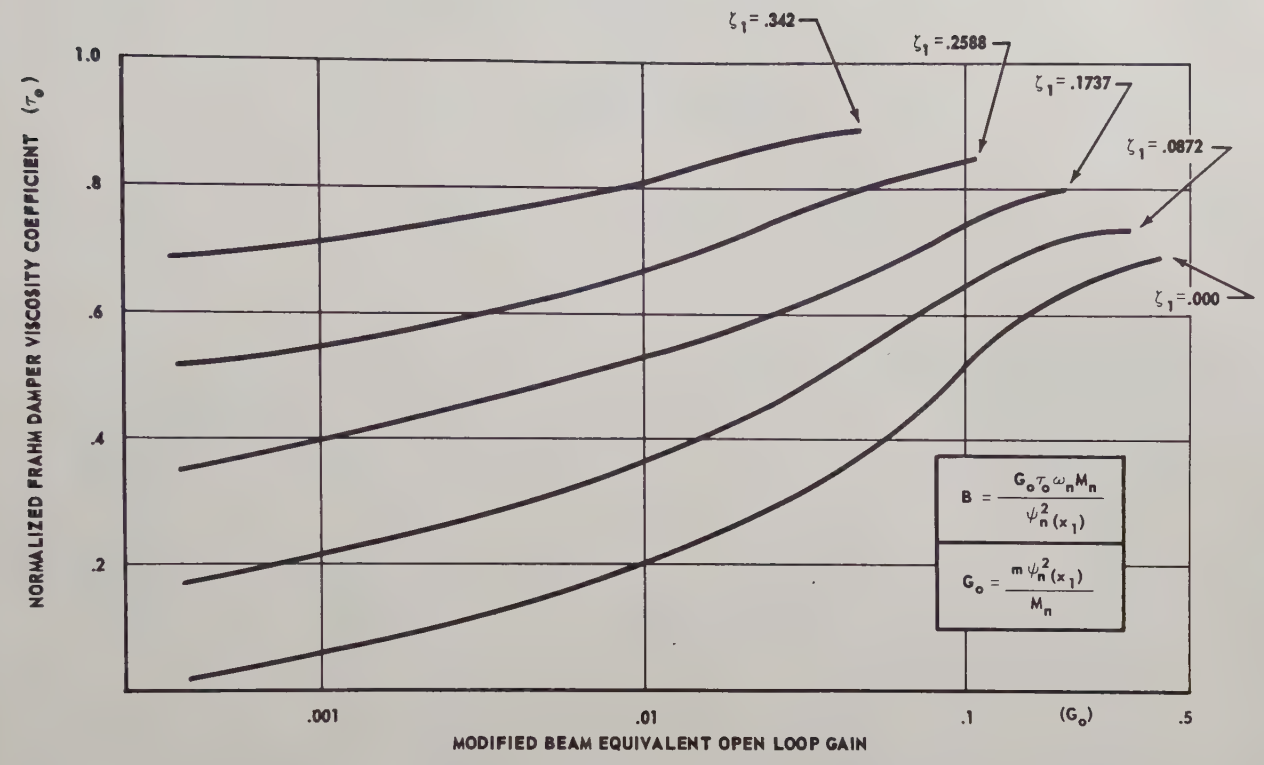


Fig. 6. Dashpot coefficient (B) vs open loop gain (G_0) and initial beam damping (ζ_1).

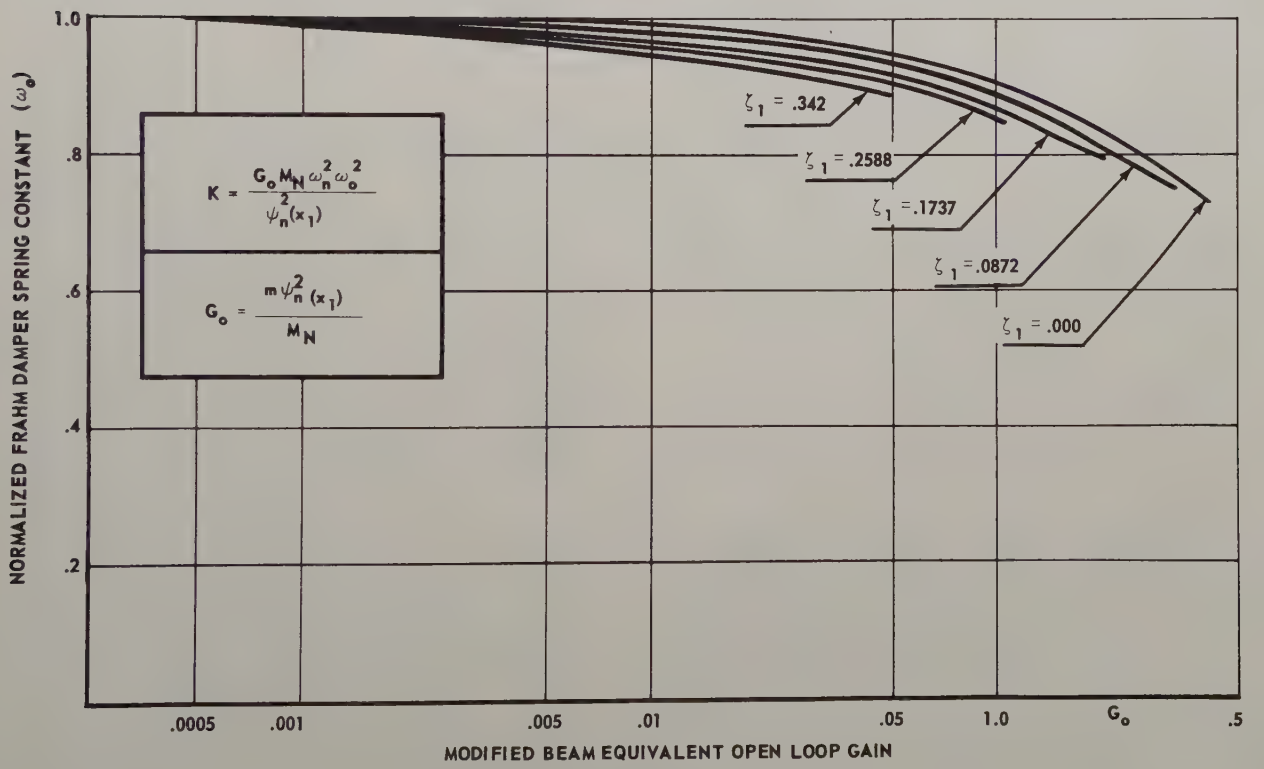


Fig. 7. Frahm spring constant (K) vs open loop gain (G_0) and initial beam damping (ζ).

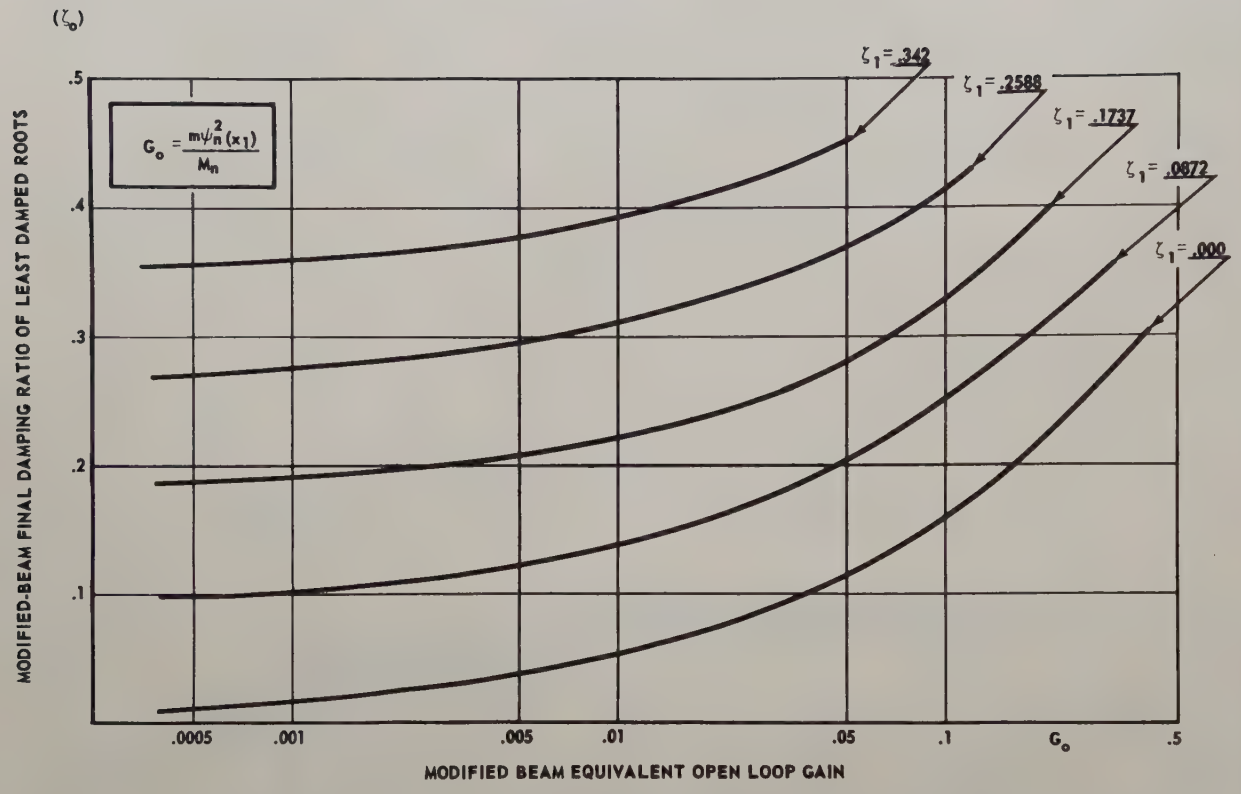


Fig. 8. Modified beam final ratio (ζ_o) vs equivalent beam open loop gain (G_o) and initial beam damping ratio (ζ).

FLYWHEEL CONTROL OF SPACE VEHICLES

by

James E. Vaeth
The Martin Company
Baltimore, Maryland

Presented at the IRE National Convention

March 22, 1960

New York, New York

Summary

An analytic study of flywheel autopilots for attitude control of space vehicles is described, and the results of a three-axis analog computer program are presented.

The objectives of this study were to determine the requirements, capabilities and limitations of flywheel autopilots as functions of desired accuracy and speed of response, disturbing moments, differential gravity restoring torques, component uncertainties, and vehicle initial attitude and attitude rate errors. The stability and accuracy degradations that resulted from not compensating for gyroscopic crosscoupling torques are summarized. Design tools are presented for synthesizing a three-axis autopilot in accordance with specified design criteria of size, power, accuracy, response rate and error disturbances.

Introduction

The use of accelerating flywheels for vehicle attitude control has been considered and proposed for various space missions, particularly when long duration flights and accurate vehicle orientation are required.[1-3]

The basic equations relating vehicle response to flywheel reaction torques and various disturbing moments have been covered in considerable detail. However, the actual synthesizing of a three-axis autopilot system based on various design criteria is usually neglected. When flywheel controllers are used, the effects of intermode coupling between the pitch, roll and yaw axes are especially significant.

This paper presents the results of a study program which was undertaken to provide a better understanding of the system synthesis problem, including the effects of intermode coupling.

Control Application

The ability of inertia wheels to control vehicle attitude stems from the basic law of action and reaction. If an inertia wheel mounted within the vehicle is accelerated about its axis of rotation, the vehicle is accelerated in inertial space about the same axis, but in the opposite direction. Assuming zero initial conditions,

$$\ddot{\theta} = -I\dot{\sigma}/P \quad \text{(refer to list of symbols).} \qquad (1)$$

Utilization of this basic principle to simultaneously control vehicle attitude about all three axes requires an autopilot consisting of attitude displacement sensors and/or rate sensors, control wheels, and electronic circuitry which must govern the rotation of the wheels as a function of the attitude displacement and/or rate signals. Figure 1 presents a functional block diagram of a single-axis control loop. Three such loops are ordinarily required, one for each axis. The attitude displacement sensor could be a gyroscope or an inertial platform for short duration missions, but in long duration missions (for which flywheels are especially attractive) gyro drift rate would have to be corrected by means of some monitoring device such as an horizon scanner or a stellar tracker. Aerodynamic forces are negligible when traveling in space and, consequently, attitude damping must be artificially provided by use of rate signals, either from a rate gyro as shown in Fig. 1 or from a lead circuit operating on the displacement error signal. Flywheel acceleration, in response to the sensor error signals, provides the restoring torque necessary to control vehicle attitude in the presence of initial condition errors and disturbing moments. Gyroscopic torques about each axis result from rotations of the vehicle and of internal machinery about the other two axes.

From the viewpoint of system weight and complexity, the differential gravity torque shown in Fig. 1 provides an attractive method for stabilizing an earth satellite to the local vertical (in a low altitude near-circular orbit). This method will be subsequently discussed in some detail.

The coordinate reference frame employed and the desired orientation of the vehicle are important considerations. Figure 1 is applicable with either an earth-fixed local vertical system or an

inertial frame fixed with respect to the stars. The former is desirable for a communication or reconnaissance satellite, whereas the latter is more advantageous for an astronomical satellite or an interplanetary space probe. If a local vertical orientation is maintained, additional crosscoupling terms are introduced as a result of the rotation of the coordinate reference frame in inertial space.

Study Objectives and Approach

In order to evaluate the requirements, capabilities and limitations of flywheel controllers, an analytic program with the following objectives was undertaken.

(1) What is the most efficient form of the controller transfer function (ratio of flywheel angular velocity to sensed error signals)?

(2) What compromises are possible as functions of attitude accuracy requirements, speed of response, initial condition and component inaccuracies, disturbing moments, wheel saturation, and system size, weight and required power?

(3) What stability and accuracy degradations result from the gyroscopic crosscoupling torques?

(4) What is the best way to compensate for these crosscoupling moments and what complexity is involved?

The analytic study was accomplished by means of a two-phase program. Phase I involved single-axis hand analyses for approximately determining the first two objectives, whereas a three-axis analog computer program was undertaken to evaluate the crosscoupling effects.

In view of the study objectives noted, no attempt was made to determine an optimized attitude displacement error sensor. Depending upon the specific space mission, various types would be applicable and it was presumed that any necessary coordinate conversions could be accomplished with zero error. The investigation concentrated on two types of attitude reference systems. The first assumed a desired orientation fixed in inertial space and a continuous source of three-axis, attitude displacement error signals within the space vehicle. The second assumed a coordinate reference frame defined by the earth's local vertical and orbital plane (that is, rotating in inertial space) but with no source of continuous displacement signals available in the vehicle. The latter would utilize differential gravity as a restoring torque, with rate gyros and flywheels for attitude damping. For this orientation, the programmed orbital angular velocity would have to be subtracted from the rate gyro output signal.

Single-Axis Investigation

The optimized form of the controller transfer function was determined by studying the transient response of the system to various forcing functions and initial conditions. Referring to Fig. 1 and assuming a displacement error sensor, the idealized controller transfer function is

$$\frac{\sigma}{\epsilon_1} = \frac{-1}{IS} \tag{2}$$

The integration 1/S is needed to prevent the attitude error (θ_ϵ) from building up due to a small but constant disturbing torque (M_E). For example, if σ/ϵ_1 has the form -1/I instead of -1/IS, then

$$\frac{\theta_\epsilon}{M_E} = \frac{-1}{K_1 S \left[1 + \frac{(K_2 + P)}{K_1} S\right]} \tag{3}$$

Another advantage of using the controller defined by Equation (2) is its suitability to crosscoupling torque compensation. For example, if the gyroscopic moments are measured and inserted as compensating signals into the controller (Fig. 1), the resulting control moment (M_C) would exactly equalize the disturbing torque.

The characteristic equation for the one-axis attitude control loop, assuming Equation (2) and a displacement error sensor, is

$$1 + \frac{K_2}{K_1} S + \frac{P}{K_1} S^2 = 0. \tag{4}$$

Speed of response and accuracy are both improved as K_1 is increased and system damping is governed by K_2. For example, a damping ratio of 0.5 is attained by making

$$K_2 = (K_1 P)^{1/2}. \tag{5}$$

Transfer functions of system response to various forcing functions are shown in Table 1, and transient response plots (for example, Fig. 2) for each equation in Table 1 were determined for various values of the significant parameters. Table 1 indicates that the steady-state value of

θ_ϵ that would result from a constant disturbing moment or rate gyro drift would vary inversely with displacement gain (K_1). The primary problem associated with this control scheme is that the flywheel continuously accelerates in the presence of a constant disturbing moment (as noted from Equation (4) of Table 1). There appears to be no controller form which can remedy this situation, unless some additional restoring torque (such as differential gravity, solar sail or aerodynamic force) is introduced. Consequently, these disturbing torques, although small, would have to be equalized for this particular type of system by fixed reaction jet impulses which would be fired when the flywheel approaches saturation.

The use of differential gravity to remedy flywheel saturation was investigated. For this type of system which utilizes only a rate gyro and flywheel controller, the optimum form of the controller was determined to be

$$\frac{\sigma}{\epsilon_1} = \frac{-K_0}{1 + \tau S}. \qquad (5)$$

For this case, a pure integration such as $\frac{-K_0}{S}$, although more desirable from the viewpoint of system simplicity and stability, would cause the flywheel to continuously accelerate due to a rate gyro drift or an error in the preprogrammed orbital rate.

The characteristic equation, using Equation (5) with $K_0 = \tau / I$, is

$$1 + \tau S + \frac{(P + \tau K_2)}{K_4} S^2 + \frac{\tau P}{K_4} S^3 = 0. \qquad (6)$$

where the differential gravity restoring torque, as derived for the small angular displacements from the local vertical [4], is defined for the pitch (y) axis as

$$K_4 = 3\omega_0^2 (P_x - P_z). \qquad (7)$$

Redrawing Fig. 1 as Fig. 3 simplifies the system synthesis problem, which is essentially the selection of optimum values of τ and K_2. P and K_4 are fixed by the vehicle configuration and orbital period (a near-circular orbit is required for this type of system to be effective). Referring to Fig. 3, satisfactory system damping is provided by making

$$\tau^2 \cong \frac{P + \tau K_2}{K_4} \quad \text{and} \quad \tau K_2 \geq P.$$

Transfer functions defining the response of such a system to various forcing functions are shown in Table 2, and Fig. 4 shows a typical transient response plot of vehicle attitude error and flywheel momentum. Table 2 indicates that the flywheel will not continuously accelerate in the presence of a steady-state disturbing torque (Fig. 4) or rate gyro drift. However, because of the small magnitude of K_4, as defined by Equation (7), system accuracy and speed of response are, of necessity, quite low. A consequent limitation to the use of this type system is that initial conditions (in particular vehicle angular rate) must be fairly precise. As will be subsequently described, this problem is increased by an order of magnitude for the practical case of simultaneous three-axis control.

Three-Dimensional Equations of Motion

The three-dimensional equations of motion for the case in which the desired attitude orientation remains fixed in inertial space, as derived in Appendix A, are presented in Table 3. It is noteworthy that the preceding single-axis analyses neglected the final three terms on the left side of each equation. These terms represent the gyroscopic crosscoupling torques--the product of angular momentum about one axis multiplied by angular velocity about the second axis produces a torque about the third axis.

For the desired orientation defined by the earth's local vertical and orbital plane, the system equations become considerably more complex. However, by assuming small angular deviations from the desired vehicle orientation and a constant orbital angular rate (ω_0), the equations can be managed. These are derived by making the following substitutions into the equation of Table 3.

$$\begin{aligned} \dot{\theta} &= \dot{\theta}_L - \omega_0 & \ddot{\theta} &= \ddot{\theta}_L \\ \dot{\phi} &= \dot{\phi}_L - \omega_0 \psi_L & \ddot{\phi} &= \ddot{\phi}_L - \omega_0 \dot{\psi}_L \\ \dot{\psi} &= \dot{\psi}_L + \omega_0 \phi_L & \ddot{\psi} &= \ddot{\psi}_L + \omega_0 \dot{\phi}_L \end{aligned} \qquad (8)$$

A further simplification ($\sigma_x = 0$) is attained by using a minimum equipment system [2] employing two flywheels and two rate gyros. The rate gyros sense angular velocities about the vehicle pitch and roll principal inertia axes, whereas the flywheels are aligned along the pitch and yaw axes--the pitch axis is defined as being normal to the orbital plane and the yaw axis through the center of the earth for the case of zero attitude error. With this configuration, roll control re-

sults from gyroscopic coupling (through ω_0) with the yaw control mode which does use a flywheel. The differential gravity restoring torques, which must be added to the equations of Table 3, are as follows.

$$M_{E_x} = 3\omega_0^2 (P_y - P_z) \phi_L$$

$$M_{E_y} = 3\omega_0^2 (P_x - P_z) \theta_L$$

$$M_{E_z} = 0 \qquad (9)$$

Three-Dimensional Simulation Program

The gyroscopic crosscoupling effects were investigated by mechanizing the three equations of motion in The Martin Company analog computer facility. This program utilized the controller transfer functions which were optimized by means of the single-axis synthesis, with allowances for gain and time constant variations. Provisions were made for varying all of the parameters shown in Fig. 1 and for introducing various inputs into each of the three control loops, either individually or simultaneously.

The gyroscopic moments, computed in accordance with the equations in Table 3, were introduced as disturbing moments and also, in some cases, as compensating signals to the flywheel controllers. In addition to constant disturbing torques, impulse torques were introduced to simulate a meteoritic impact, an initial velocity error or a change in angular momentum associated with vehicle rotating equipment.

A summation of the significant results of this simulation program follows. In addition to the analog study, the roots of the roll-yaw characteristic equations were determined by digital computation for the system employing differential gravity and two wheels. These results showed excellent correlation with the response plots from the three-axis analog program.

Computer Results of the Three-Wheel System

The three-axis analog simulation of the system utilizing an inertial frame of reference and three flywheels fixed to the vehicle principal axes revealed little or no stability degradation due to the crosscoupling terms, with or without compensation. However, the gyroscopic moments (when not compensated) caused significant transient errors in vehicle attitude, but in all cases these errors subsequently damped to zero. System response to attitude displacement commands, disturbing torques and rate gyro errors was determined for a wide variation of gain settings--system period of oscillation ranged from 60 to 2800 seconds for all axes. The transient response plots were as expected from the single axis analyses (Table 1).

A typical transient response plot is shown in Fig. 5, for which the following set of parameters was used.

$$P_x = P_y = 1000 \text{ slug-ft}^2$$

$$P_z = 100 \text{ slug-ft}^2$$

Controller (each axis): $\frac{\sigma}{\epsilon_1} = \frac{-1}{IS\,(1 + 0.5\,S)}$

K_1 was fixed so that the approximate natural frequency was 0.1 radian/second for each axis.

K_2 was fixed so that the approximate damping ratio was 0.5 for all axes.

Figure 5 plots system response to two types of inputs. The first, at time t = 100 seconds, was a pitch impulse disturbance of 11 foot-pound-seconds (equivalent to a vehicle initial angular rate of 0.63 degree/second in inertial space) and assumed net angular momentums of zero about the roll and yaw axes. As expected from the single-axis studies (Table 1), the pitch flywheel absorbs this momentum change after which pitch attitude is stabilized to zero error. From this and similar transient plots, a good approximation for determining peak overshoot, assuming a 0.5 damping ratio, is

$$(\theta_\epsilon)_{\text{peak}} \simeq \frac{-\Delta H}{2(PK_1)^{1/2}} \qquad (10)$$

where $\Delta H = M_{D/S}$ or $\Delta H = P\dot{\theta}_i$. If the damping ratio is decreased by using a lower value of K_2, $(\theta_\epsilon)_{\text{peak}}$ would be increased.

System response to the second input, a roll displacement command of 10 degrees, illustrates the gyroscopic crosscoupling effects. The peak yaw displacement of 0.084 radian in Fig. 5 results from the induced yawing moment defined by $\dot{\phi}\, I_y\, \sigma_y$. From Fig. 5, $I_y\, \sigma_y$ = 11 foot-pound-seconds and the average $\dot{\phi}$ for a 30-second time interval is approximately 0.007 radian/second. Substituting these values into Equation (10), $(\psi_\epsilon)_{\text{peak}}$ is determined to be 0.082 radian, which

explains the observed results. Gyroscopic coupling into the pitch axis is seen to be considerably less because $(P_y K_1)^{1/2}$ (Equation (10)) is greater by an order of magnitude than $(P_z K_1)^{1/2}$. The residual yaw wheel momentum of 2 ft-lb-sec results from the gyroscopic torque $(\dot{\phi} \sigma_y I_y)$ applied for 30 seconds in one direction. This increment of yaw momentum was absorbed by the yaw flywheel. When roll displacement was commanded back to zero, yaw wheel momentum returned to zero.

When gyroscopic torque compensation was applied, the response of all three flywheels, as plotted in Fig. 5, remained essentially the same. However, the yaw attitude excursions were eliminated by introducing these compensating signals (the final three terms on the left side of each equation in Table 3) into the flywheel controllers. In other analog runs with the controller gain set at 1.25/I instead of 1.0/I, compensation was approximately 75% effective.

Simultaneous attitude displacement commands of 10 degrees about each axis were also introduced. Response was quite satisfactory even without gyroscopic torque compensation; some degradation in transient response resulted, particularly when an initial angular momentum was present, but all axes stabilized to zero within the expected time duration. A typical response for a low gain system is reproduced in Fig. 6. The noticeable lag in yaw response is attributed to the comparatively low vehicle inertia about the yaw axis together with the gyroscopic torque $(\dot{\phi} I_y \sigma_y)$. With this inertial frame of reference, the recommended vehicle configuration would have $P_x = P_y = P_z$. This inertia ratio is almost mandatory in a low-orbit earth satellite because of differential gravity torques. The alternative would be to utilize differential gravity to help provide a local vertical orientation with $P_y > P_x >> P_z$.[4]

Computer Results of the Two-Wheel System

Initial analog results of the two-wheel system, using differential gravity together with pitch and yaw controller transfer functions as defined by Equation (5), indicated marginal stability. A roll-yaw damping ratio of approximately 0.15 was the best that could be obtained by varying τ and K_2, whereas pitch response was as expected from the single-axis analyses (Fig. 4). Because of this roll-yaw stability degradation and the increased complexity introduced into the analog simulation by Equations (8) and (9), the analog program was terminated after the initial runs to conduct additional analysis. The three-axis system, as simulated, was reduced to a two-axis roll-yaw coupled mode by assuming σ_y and $\dot{\theta}$ constant. The roots of the resulting sixth order characteristic equation were extracted and, in all cases, good correlation (in terms of frequency of oscillation and speed of convergence or divergence) was established with the analog response plots. The correlation demonstrated the effectiveness of this root extraction method for system optimization and also substantiated that a 0.15 damping ratio was all that could be expected.

The form of the roll-yaw characteristic equation is developed as Equation (B-3) in Appendix B. Tables 4 and 5 illustrate the resulting roots for some typical cases and also show the criticality of τ and K_2. Varying these values by a factor of two can change an optimum damped system to one of negligible damping or actual instability.

Table 5 shows that stability need not be decreased by incorporating a constant wheel momentum on the pitch axis, although fluctuations of the yaw control parameters τ and K_2 (which must be selected for this constant value of $I_y \sigma_y$) appear to be even more critical. This increased pitch momentum, having the same polarity as ω_0, provides additional yaw stiffness and thereby reduces the yaw transient error resulting from an initial yaw or roll rate or a subsequent disturbing torque. The analog response plots verified that the peak overshoot in yaw could be approximated from Equation (10). The actual peak values were somewhat greater due to damping of less than 0.5. In applying this equation, K_1 must be replaced by the differential gravity restoring torque (K_4) about the yaw axis, which can be approximated (see Appendix B) as

$$\omega_0^2 (P_y - P_x - I'_y)$$

where $\omega_0 I'_y$ is equivalent to the constant value of pitch wheel momentum $(I_y \sigma_y)$. Thus, with a constant $I_y \sigma_y$ of -0.9 ft-lb-sec as assumed in Table 5, it follows from Equation (10) that the peak yaw overshoot would be decreased by 2.24 as compared to the case with zero I'_y.

Physical Characteristics

The physical requirements of the system in

terms of flywheel size, weight and power depend upon the maximum wheel momentum needed together with its rate of change. From Equation (1), the wheel momentum must be sufficient to provide a specified maximum slewing rate of the vehicle. However, if vehicle stabilization alone is required, the wheel need only equalize disturbing torques and initial conditions. For this situation, the flywheel physical requirements can be held to a minimum by compromising system speed of response. K_1 should be just sufficient to provide the specified static accuracy. The equations in Tables 1 and 2 together with accurate predictions of initial conditions and disturbing torques provide the basis for selecting minimum size components consistent with the vehicle configuration and desired performance.

A parametric plot of flywheel size and weight is presented in Fig. 7 as a function of required angular momentum. Figure 8 shows required power as a function of wheel momentum and its rate of change where

$$\text{Watts (average)} = \frac{1.36\, I\, (\Delta\sigma)^2}{2t}$$

$$\text{or } \frac{7.5\, I\, (\Delta\text{rpm})^2}{10^3 t} \qquad (12)$$

and t is the time required to accomplish a $\Delta\sigma$ (radians/second) change in wheel velocity. Minimum power is obtained by using a maximum inertial wheel, as indicated from Equation (12).

Typical values of response time and flywheel momentum are shown in Figs. 2, 4, 5 and 6. These values illustrate the significance of system gain and vehicle inertia in determining required physical characteristics.

The recommended controller transfer function defining the ratio of flywheel response to sensed error signals ($\sigma/\epsilon_1 = -1/IS$) can be realized using off-the-shelf components. A direct-drive motor and electronic circuitry incorporating a single integration would suffice for each axis. Circuit design and repeatability are not critical. A controller smoothing time of as much as 0.5 second and a 25% gain variation were shown to be tolerable.

If dynamic accuracy requirements necessitate compensation for gyroscopic moments, rate generators fixed to each flywheel shaft and the associated computing circuitry would have to be incorporated.

Applications and Problem Areas

The single-axis analyses and computer studies previously described made use of certain linearizing assumptions (such as small angle approximations). Also, various second order disturbances (such as orbit eccentricity) were neglected. Eccentricity effects could be serious with a low gain system and, in particular, when differential gravity is used; the control system natural frequency would be of the same order of magnitude as that of the sinusoidal input disturbance.

Although more complex computer programs are certainly recommended, it is believed that such effects can reasonably be inferred from the results of this and other publications.[5, 6] For example, computer transient response plots of vehicle oscillation about the local vertical[5] for an undamped "dumbbell" configuration indicate half amplitude oscillations of 6 degrees peak for a 7% eccentric orbit. Moreover, it appears that the amplitude of this oscillation would not be substantially reduced by the flywheel damping system--computer response plots of a mechanical passive damper[6] substantiate this conclusion.

As an outgrowth of the analytic results presented, it seems pertinent to point out various space missions for which each of the two types of flywheel controllers are recommended. If precise vehicle orientation is required, such as in astronomical satellites or during periods of stellar navigation "fixes," the controller form defined by Equation (2) should be employed. This permits the use of a relatively high gain system to minimize errors in accordance with the equations of Table 1. It also provides a low-torque linear control capability. However, for each particular application requiring precise accuracy, the actual controller hardware should be breadboarded and included in a three-axis analog computer study to evaluate the effects of any existing nonlinearities. The predominant drawback is the tendency of the flywheels to saturate. This problem could be remedied by using fixed low-thrust reaction jets to desaturate the wheels (such jets would probably be present for initial angular rate stabilization). If this controller is employed to provide a local vertical orientation, such as with a high-accuracy reconnaissance satellite, the relationships in Equations (8) and (9) would have to be included in the three-axis equations of motion as presented in Table 3. Although this degree of sophistication was not achieved in the analog program, no stability or accuracy degradations are to be expected unless an extremely low displacement gain is used.

If attitude deviations from the local vertical of between 1 and 10 degrees are permissible, as might be the case with a communication or meteorological satellite, the minimum equipment system utilizing differential gravity would have application. The recommended form of the pitch and yaw controllers is as defined in Equation (5). As noted from Table 2, the saturation problem is circumvented, but other problems are introduced because of the very low static gain (differential gravity restoring torque). Consequently, the effective use of this type control presupposes extremely low initial angular rates and a near-circular orbit. Another attractive use of this type system, as in a reconnaissance satellite, would be as a backup in the event the airborne attitude reference system should malfunction. In addition to simplicity of operation, the desired vehicle configuration ($P_y > P_x >> P_z$) provides the possibility of indefinitely maintaining a local vertical orientation after all electronic equipment has been shut down. One other "dumbbell" application would be for crude alignment of an earth satellite, with fine alignment provided by gimbaling the reconnaissance sensors.

Conclusions

Principal results and accomplishments are summarized as follows.

(1) Three-axis flywheel control of vehicle attitude can be effectively employed. The low-torque linear control characteristics provide an excellent accuracy capability.

(2) When the desired vehicle orientation is fixed in inertial space, gyroscopic coupling has negligible effect on system stability. Compensation for such effects can significantly improve dynamic accuracy, but the additional complexity is not warranted for most missions.

(3) With a desired local vertical orientation and a low response control system (that is, differential gravity), stability is significantly affected by gyroscopic coupling.

(4) Effective design tools have been developed, with illustrative examples, for synthesizing a three-axis flywheel autopilot as functions of desired accuracy and speed of response, vehicle inertia characteristics, disturbing torques, component inaccuracies, vehicle initial attitude and attitude rate errors, and flywheel size, weight and power requirements.

(5) Major problem areas are system reliability during long missions and flywheel saturation.

(6) With a low-altitude near-circular orbit, the use of differential gravity minimizes both of these problems, provided precise initial conditions can be achieved.

References

1. "Preliminary Considerations on the Instrumentation of a Photographic Reconnaissance of Mars," Laning, Frey and Trageser, Instrumentation Laboratory, Massachusetts Institute of Technology, April 1958.

2. "Attitude Control--II," Roberson, R.E., AIEE Space Vehicle Symposium, Buffalo, New York, June 1958.

3. "An Orbiting Astronomical Telescope Design Study," Schroader and Chubbuck, 14th Annual Meeting of the American Rocket Society, Washington, D.C., 16 to 20 November 1959.

4. "Determination of an Unique Attitude for an Earth Satellite," Davis, W.R., 4th Annual Meeting of American Astronautical Society, New York, New York, 30 January 1958.

5. "The Two-Dimensional Librations of a Dumbbell-Shaped Satellite in a Uniform Gravitational Field," Stocker and Vachino, Air Force Institute, March 1958.

6. "Passive Damping of the Two-Dimensional Librations of a Dumbbell-Shaped Satellite," Hall and Smith, Air Force Institute, March 1959.

List of Symbols

I	Flywheel moment of inertia
σ	Flywheel angular velocity
P	Moment of inertia about principal inertia axis of vehicle
x, y, z	Principal inertia axes of vehicle, assumed orthagonal in this study
ϕ	Vehicle rotation (in inertial space) about x axis
θ	Vehicle rotation (in inertial space) about y axis
ψ	Vehicle rotation (in inertial space) about z axis
M_C	Flywheel reaction torque acting upon vehicle
M_E	External torque acting upon vehicle
M_T	Total moment acting upon vehicle
H	Angular momentum
S	Laplace operator
K	Gain factor
τ	Time constant
t	Time or time interval
X	Vector cross product

ω_0	Orbital angular velocity
ω_V	Vehicle angular velocity in inertial space
I'_y	Equivalent value of constant wheel inertia along the y axis
ϵ_1	Error signal to flywheel controller
α	Attenuation factor
Δ	Increment of
rpm	Revolutions per minute
G	Rate gyro inaccuracy or error in preprogrammed angular velocity of vehicle
C(S)	System characteristic equation
R	Flywheel radius of gyration

Subscripts

L	Refers to vehicle motion with respect to local vertical orbital-plane coordinate system defined (for zero error) as having the z axis through the earth's center and the y axis normal to the orbital plane.
F	Refers to flywheels
V	Refers to vehicle
c	Refers to command (or zero reference) attitude
ϵ	Refers to attitude error
i	Refers to initial condition
u	Refers to unit vector
x, y, z	Refers to principal inertia axes of vehicle

Superscripts

$\underline{\quad}$	Vector quantity
.	d/dt
..	d^2/dt^2

TABLE 1

One-Axis Transfer Functions with Inertial Coordinate System

$$\frac{\theta}{\theta_c} = \frac{1}{C(S)} \tag{1}$$

$$\frac{\theta_\epsilon}{M_E} = \frac{\theta_\epsilon}{\dot{\theta}_i PS} = \frac{-1}{K_1 C(S)} \tag{2}$$

$$\frac{\theta_\epsilon}{G} = \frac{K_2}{K_1 C(S)} \tag{3}$$

$$\frac{I\sigma}{M_E} = \frac{I\sigma}{\dot{\theta}_i PS} = \frac{(1 + K_2/K_1 S)}{S C(S)} \tag{4}$$

$$\frac{I\sigma}{G} = \frac{K_2 PS}{K_1 C(S)} \tag{5}$$

Controller Form:

$$\frac{\sigma}{\epsilon_1} = \frac{-1}{IS}$$

$$C(S) = 1 + \frac{K_2}{K_1} S + \frac{P}{K_1} S^2$$

TABLE 2

Pitch-Axis Transfer Functions with Differential Gravity

$$\frac{\theta_\epsilon}{M_E} = \frac{\theta_\epsilon}{\dot{\theta}_{L_i} P_y S} = \frac{-(1 + \tau S)}{K_4 C(S)} \tag{1}$$

$$\frac{\theta_\epsilon}{G} = \frac{-K_2 \tau S}{K_4 C(S)} \tag{2}$$

$$\frac{I_y \sigma_y}{M_E} = \frac{I_y \sigma_y}{\dot{\theta}_{L_i} P_y S} = \frac{K_2 \tau S}{K_4 C(S)} \tag{3}$$

$$\frac{I_y \sigma_y}{G} = \frac{-\tau K_2 (1 + P_y/K_4 S^2)}{C(S)} \tag{4}$$

Controller Form:

$$\frac{\sigma_y}{\epsilon_1} = \frac{-\tau}{I_y (1 + \tau S)} \quad \text{where } \tau \geq \left(\frac{P_y}{K_4}\right)^{1/2}$$

$$K_4 \tilde{=} 3\omega_0^2 (P_x - P_z)$$

$$C(S) = 1 + \tau S + \frac{(\tau K_2 + P_y)}{K_4} S^2 + \frac{\tau P_y}{K_4} S^3$$

TABLE 3

Three-Dimensional Equations of Motion

$$I_x \dot{\sigma}_x + P_x \ddot{\phi} + \dot{\theta}\,\sigma_z I_z - \dot{\psi}\,\sigma_y I_y + \dot{\theta}\,\dot{\psi}\,(P_z - P_y) = M_{E_x} \quad \text{(roll axis)}$$

$$I_y \dot{\sigma}_y + P_y \ddot{\theta} + \dot{\psi}\,\sigma_x I_x - \dot{\phi}\,\sigma_z I_z + \dot{\psi}\,\dot{\phi}\,(P_x - P_z) = M_{E_y} \quad \text{(pitch axis)}$$

$$I_z \dot{\sigma}_z + P_z \ddot{\psi} + \dot{\phi}\,\sigma_y I_y - \dot{\theta}\,\sigma_x I_x + \dot{\phi}\,\dot{\theta}\,(P_y - P_x) = M_{E_z} \quad \text{(yaw axis)}$$

I ≡ Flywheel inertia

σ ≡ Flywheel angular velocity

$\dot{\theta}, \dot{\phi}, \dot{\psi}$ ≡ Vehicle pitch, roll and yaw inertial rates

P ≡ Vehicle inertia

M_E ≡ External torques

x, y and z refer to the roll, pitch and yaw principal inertia axes.

TABLE 4

Roots of Roll-Yaw Characteristic Equation with $I'_y = 0$

τ (sec)	K_2	Roots				Minimum Damping Ratio
100	2.25	-0.0081,	-0.010,	-0.00017 ± 0.0021,	-0.0007 ± 0.0010	0.081
200	2.25	-0.00055,	-0.0050,	-0.00010 ± 0.0021,	-0.0021 ± 0.0028	0.048
400	2.25	-0.00023,	-0.0028,	+0.000062 ± 0.0019,	-0.0011 ± 0.0039	Divergent -0.033
400	4.5	-0.00011,	-0.0025,	-0.000012 ± 0.0021,	-0.0012 ± 0.0053	0.006
400	1.125	-0.00047,	-0.0025,	-0.00015 ± 0.0021,	-0.00085 ± 0.0025	0.071
200	1.125	-0.0031,	-0.0050,	-0.000094 ± 0.0022,	-0.00086 ± 0.0011	0.043
400	0.5625	-0.0011,	-0.0025,	-0.00015 ± 0.0022,	-0.00055 ± 0.0015	0.068
200	0.5625	-0.0042,	-0.0051,	-0.000076 ± 0.0022,	-0.00031 ± 0.0012	0.034
50	0.5626	-0.0189 ± 0.0045,		-0.000019 ± 0.0022,	-0.00007 ± 0.0011	0.009

$P_y = 2000$ slug-ft^2, $P_x = 1800$ slug-ft^2, $P_z = 200$ slug-ft^2

$\omega_0 = 0.0011$ rad/sec

TABLE 5

Roots of Roll-Yaw Characteristic Equation with $I'_y \omega_0 = -0.9$ slug-ft^2/sec

τ (sec)	K_2	Roots			Minimum Damping Ratio
100	1.125	-0.0092 ± 0.0029,	-0.000083 ± 0.0031,	-0.00020 ± 0.0018	0.027
200	1.125	-0.0042 ± 0.0017,	-0.00050 ± 0.0028,	-0.00033 ± 0.0022	0.15
400	1.125	-0.0017 ± 0.00062,	$+0.0000004 \pm 0.0020$,	-0.00085 ± 0.0040	Divergent -0.0002
400	0.5625	-0.00041 ± 0.0035,	-0.000040 ± 0.0019,	-0.0021 ± 0.0007	0.021
200	0.5625	-0.0049 ± 0.0017,	$+0.000033 \pm 0.0034$,	-0.00016 ± 0.0016	Divergent -0.010
200	2.25	-0.0021, -0.0050,	-0.000028 ± 0.0018,	-0.0014 ± 0.0046	0.016
800	2.25	-0.00042, -0.00125,	-0.000054 ± 0.0019,	-0.00042 ± 0.0054	0.028
800	4.5	-0.00025, -0.0019,	-0.00091 ± 0.0021,	$+0.00070 \pm 0.0046$	Divergent -0.015
400	4.5	-0.00051, -0.0025,	-0.000026 ± 0.0018,	-0.00098 ± 0.0068	0.014
200	4.5	-0.0012 ± 0.00045,	$+0.00037 \pm 0.0016$,	-0.0042 ± 0.0126	Divergent -0.23

$P_y = 2000$ slug-ft^2, $P_x = 1800$ slug-ft^2, $P_z = 200$ slug-ft^2

$\omega_0 = 0.0011$ rad/sec

APPENDIX A

THREE-DIMENSIONAL EQUATIONS OF MOTION

The moment equations for a space vehicle using three variable-speed flywheels fixed to the vehicle principal inertia axes, assuming constant vehicle and flywheel moments of inertia, are derived from the following vector equations (refer to list of symbols).

$$\bar{M}_T = \dot{\bar{H}} = \dot{\bar{H}}_F + \dot{\bar{H}}_V \text{ (assumes } M_E = 0) \qquad \text{(A-1)}$$

$$\bar{H}_F = I\bar{\sigma} \qquad \text{(A-2)}$$

$$\bar{H}_V = P\,\bar{\omega}_V \qquad \text{(A-3)}$$

$$\dot{\bar{\sigma}} = \dot{\sigma}_x \bar{x}_u + \dot{\sigma}_y \bar{y}_u + \dot{\sigma}_z \bar{z}_u + \sigma_x (\bar{\omega}_V X \bar{x}_u) + \sigma_y (\bar{\omega}_V X \bar{y}_u) + \sigma_z (\bar{\omega}_V X \bar{z}_u) \qquad \text{(A-4)}$$

$$\dot{\bar{\omega}}_V = \ddot{\phi}\,\bar{x}_u + \ddot{\theta}\bar{y}_u + \ddot{\psi}\bar{z}_u + \dot{\phi}(\bar{\omega}_V X \bar{x}_u) + \dot{\theta}(\bar{\omega}_V X \bar{y}_u) + \dot{\psi}(\bar{\omega}_V X \bar{z}_u) \qquad \text{(A-5)}$$

$$\bar{\omega}_V X \bar{x}_u = \dot{\psi}\,\bar{y}_u - \dot{\theta}\,\bar{z}_u \qquad \text{(A-6)}$$

$$\bar{\omega}_V X \bar{y}_u = \dot{\phi}\,\bar{z}_u - \dot{\psi}\,\bar{x}_u \qquad \text{(A-7)}$$

$$\bar{\omega}_V X \bar{z}_u = \dot{\theta}\,\bar{x}_u - \dot{\phi}\,\bar{y}_u \qquad \text{(A-8)}$$

Equation (A-5) assumes a coordinate axis system fixed in inertial space.

Combining the above equations according to the x, y and z components gives the complete moment equations about each of the three principal inertia axes. These equations are presented in Table 3. The external torques (M_E) would arise from such sources as atmosphere, solar radiation, mass attraction, etc.

APPENDIX B

ROLL-YAW CHARACTERISTIC EQUATION WITH ROTATING COORDINATE AXIS SYSTEM AND DIFFERENTIAL GRAVITY

The equations shown in Table 3 must be modified if the desired orientation of the vehicle is rotating in inertial space. An important application is an earth satellite, the body axes of which are to remain aligned with the local vertical. If ω_0 is assumed constant, the necessary modifications are defined by Equation (8). Also, differential gravity torques must be considered unless $P_y = P_x = P_z$. For small displacements from the desired orientation (z axis through the earth's center and y axis normal to the orbital plane), the gravity torques are defined by Equation (9).

For a "minimum equipment" system requiring only two control wheels and two rate gyros, the roll-yaw coupling (through ω_0) can be effectively analyzed by assuming constant σ_y and zero $\dot{\theta}_L$.

Using these approximations and Equations (8) and (9) together with

$$\sigma_x = 0$$

$$I_y \sigma_y = I'_y \omega_0$$

$$\sigma_z = \frac{-\tau K_2 (\dot{\phi}_L - \omega_0 \psi_L)}{I_z (1 + \tau S)},$$

the first and third equations in Table 3 reduce to

$$\phi_L \left[P_x S^2 + \frac{\omega_0 K_2 \tau}{(1 + \tau S)} S + 4 \omega_0^2 \left(P_y - P_z - \frac{I'_y}{4} \right) \right] = -\psi_L \left[\omega_0 (P_y - P_z - P_x - I'_y) S - \frac{\omega_0^2 K_2 \tau}{(1 + \tau S)} \right] \quad \text{(B-1)}$$

$$\psi_L \left[P_z S^2 + \frac{\omega_0 K_2 \tau}{(1 + \tau S)} S + \omega_0^2 (P_y - P_x - I'_y) \right] = \phi_L \left[\frac{K_2 \tau}{(1 + \tau S)} S^2 + \omega_0 (P_y - P_x - P_z - I'_y) S \right] \quad \text{(B-2)}$$

By substituting (B-2) into (B-1) and expanding, the roll-yaw characteristic equation is determined to be

$$S^6 P_x P_z \tau^2 + S^5 2 P_x P_z \tau + S^4 \left[P_x P_z + \omega_0 \tau^2 \left(\omega_0 \left\{ 4 P_z K_5 + P_x K_6 + K_7^2 \right\} + K_2 \left\{ P_x + P_z - K_7 \right\} \right) \right] + S^3 \omega_0 \tau \left[\omega_0 \left(2 P_x K_6 + 8 P_z K_5 + 2 K_7^2 \right) + K_2 (P_x + P_z - K_7) \right] + S^2 \omega_0^2 \left[\omega_0 \tau^2 \left(K_2 \left\{ K_6 + 4 K_5 - K_7 \right\} + 4 \omega_0 K_6 K_5 \right) + P_x K_6 + 4 P_z K_5 + K_7^2 \right] + S \omega_0^3 \tau \left[8 \omega_0 K_6 K_5 + K_2 (K_6 + 4 K_5 - K_7) \right] + 4 \omega_0^4 K_6 K_5 = 0 \quad \text{(B-3)}$$

where

$$K_5 = P_y - P_z - \frac{I'_y}{4}$$

$$K_6 = P_y - P_x - I'_y$$

$$K_7 = P_y - P_x - P_z - I'_y .$$

This sixth order equation can be used as a design tool for optimizing K_2, τ and I'_y. These are the critical parameters since the others are fixed by the chosen orbit and vehicle configuration. Again it must be emphasized that the orbit should be near circular and that $P_y > P_x >> P_z$ for optimum performance. A straightforward method of determining system stability involves the substitution of parametric values into Equation (B-3) and subsequent root extraction or application of Routh's stability criterion. Although this is a tedious procedure, it was accomplished for selected values of K_2, τ and I'_y in order to evaluate their significance. Tables 4 and 5 present the roots, as extracted by digital computation, for various combinations of these critical parameters. These roots demonstrate that the synthesis tools developed during the single-axis hand analyses are helpful for "ballpark" selection of parameter values, but that system stability should be verified by the use of Equation (B-3).

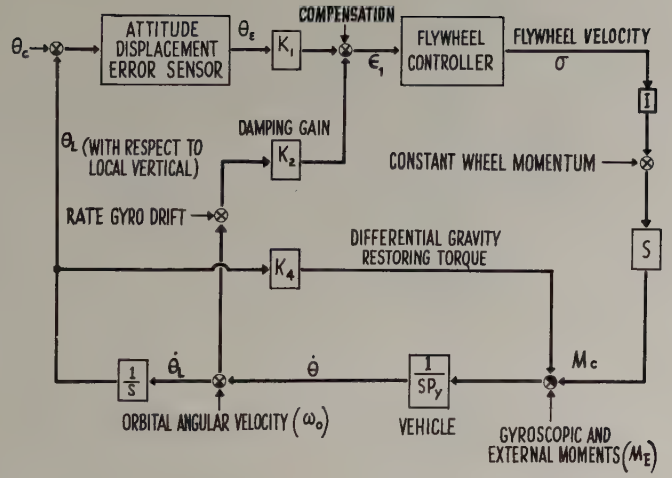
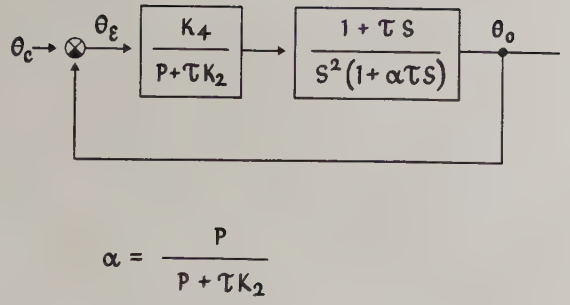


Fig. 1. Flywheel control schematic (pitch axis).



$$\alpha = \frac{P}{P + \tau K_2}$$

Fig. 3. Equivalent block diagram with differential gravity.

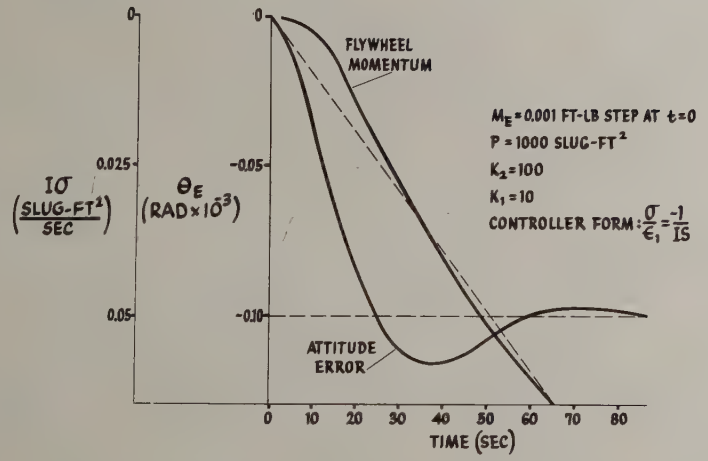


Fig. 2. One-axis transient response (inertial reference).

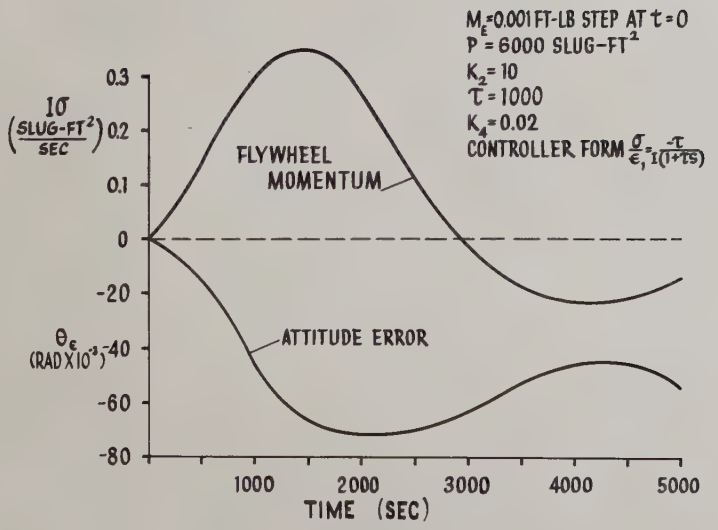


Fig. 4. Transient response with differential gravity.

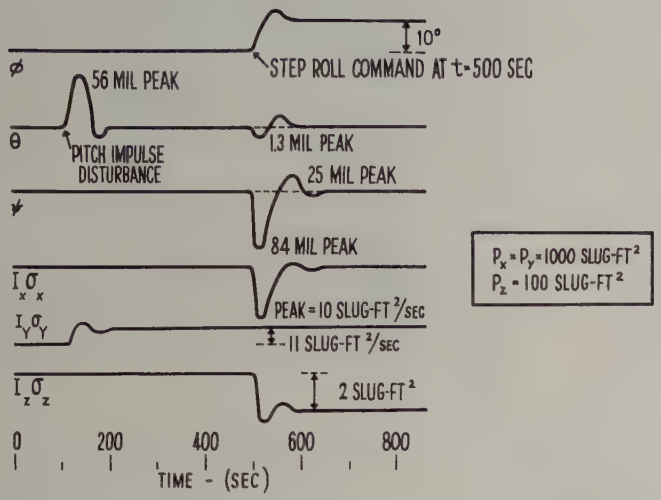


Fig. 5. Three-axis response (high gain).

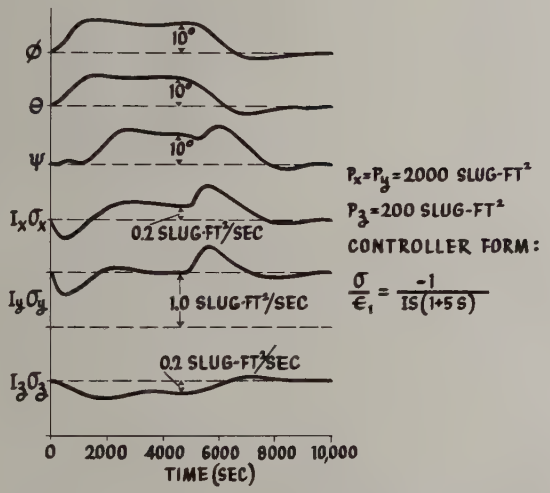


Fig. 6. Three-axis response (low gain).

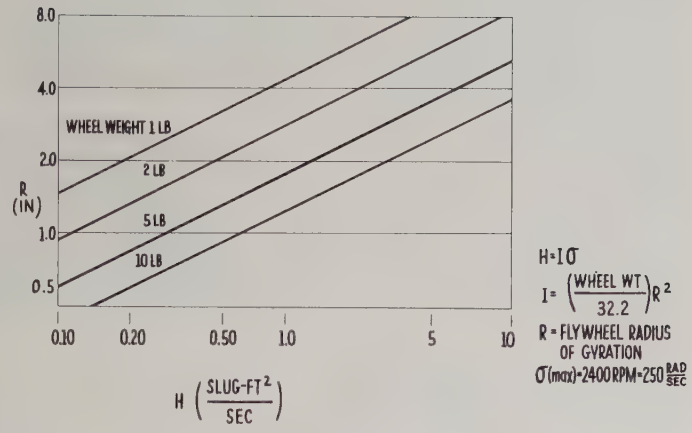


Fig. 7. Wheel size and weight vs momentum.

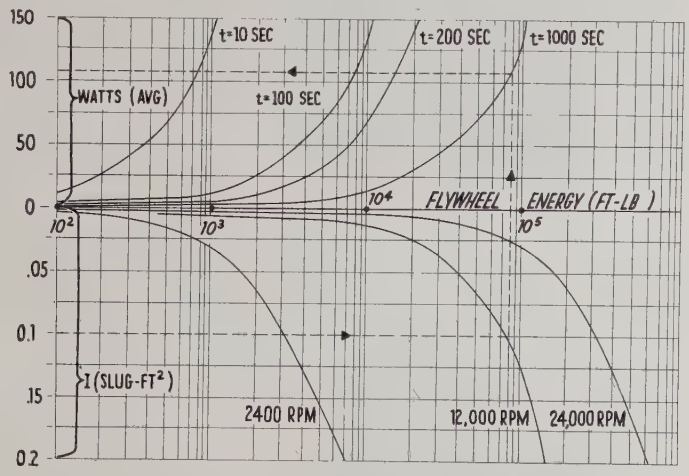


Fig. 8. Flywheel energy and power.

Contributors

Masanao Aoki (S'58-M'60) was born on May 14, 1931, in Hiroshima, Japan. He received the B.S. degree in physics in 1953 and the M.S. degree in physics in 1955, both from the University of Tokyo. He worked as a research engineer on analog to digital converters and analog computers during 1955-1956 at the Central Research Laboratory, Hitachi Ltd., in Tokyo. Since 1957 he has been at the University of California, Los Angeles, where he received the Ph. D. degree in engineering in 1960. He is presently working as a research engineer in the Department of Engineering at UCLA. Control and digital computers are his main fields of interest.

M. AOKI

. . .

John A. Aseltine (S'50-A'52-M'57) was born in Palo Alto, Calif., on April 12, 1925. He received the B.A. degree in physics in 1947 from the University of California and the Ph.D. degree in engineering in 1952 from UCLA.

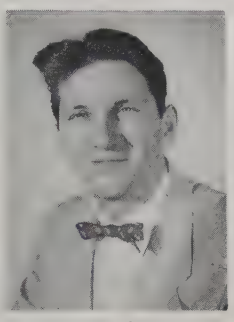
J. A. ASELTINE

During World War II he served with the Navy, and during 1947-1948 he was associated with the Radio Laboratory at Consolidated-Vultee in San Diego, Calif. From 1952 to 1954 he was a member of the technical staff at Hughes Aircraft Company, Culver City, Calif., before joining Lear, Inc., in Santa Monica. He is presently manager of the Electronics Planning and Analysis Department at Space Technology Laboratories, Los Angeles, and is a lecturer in engineering at UCLA.

Dr. Aseltine is a member of Pi Mu Epsilon and Sigma Xi.

. . .

Olle I. Elgerd (SM'59) was born in Sweden on March 31, 1925. He received the B.S.E.E. degree in 1945 from Orebo Tekniska Gymnasium, Orebo, Sweden. After serving for twelve months in the Swedish Army, he attended the Royal Institute of Technology in Stockholm and received the civ. ing. degree in 1950.



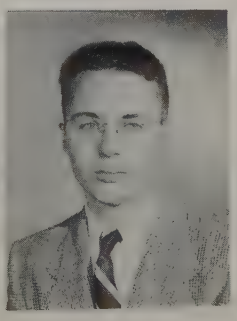
O. I. ELGERD

Before coming to the United States in 1952, he held various industrial positions in Sweden, working on the application of magnetic amplifiers to industrial and defense control problems, design and planning of hydro power plants, and design and construction of substations and transmission lines. During 1952-1953, he was with Sverdrup and Parcel, consulting engineers, in St. Louis, Mo. He was an instructor in the Department of Electrical Engineering at Washington University, St. Louis, from 1953 to 1956, while completing graduate studies for the D.Sc. degree in electrical engineering. Since 1956, he has been an associate professor at the University of Florida, Gainesville, teaching control system theory and computer techniques.

Dr. Elgerd is a member of the Swedish Association of Engineers and Architects, AIEE, and Sigma Xi. He became an American citizen in 1959.

. . .

Ronald J. Hruby (M'56) was born on October 21, 1930, in Los Angeles, Calif. He attended California Institute of Technology in Pasadena, and received the B.S.E.E. degree in 1953 and the M.S.E.E. degree in 1956. He is continuing graduate work towards the Ph.D. degree at the University of Southern California.

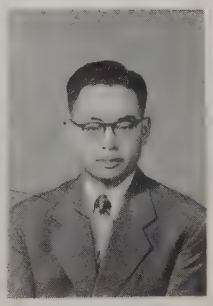


R. J. HRUBY

Mr. Hruby spent two years as an electronic engineer at White Sands Proving Ground, N. M., where he worked on the Nike and Corporal missiles. He was a senior engineer at Douglas Aircraft Company, assigned to work on control systems for the Thor and Sparrow II, and for the past year has been a senior engineer at NORAIR Division of Northrop Aircraft Company, Hawthorne, Calif., engaged in design of outer space control and data processing systems.

. . .

H. C. Hsieh was born in Foochow, China. He received the Bachelor of Science degree in electrical engineering from the National Taiwan University, Taipei, China, in 1954, and the Master of Science degree in engineering from the University of California, Los Angeles, in 1958.



H. C. HSIEH

At present he is a research engineer at UCLA, working on the adaptive control systems theory project for the Air Force Office of Scientific Research.

. . .

Cornelius T. Leondes (S'48-A'53-SM'57) was born on July 21, 1927, in Pennsylvania. He received the B.S., M.S., and Ph.D. degrees from the University of Pennsylvania, Philadelphia, in 1949, 1951, and 1954, respectively.

C. T. LEONDES

He worked briefly at the Naval Air Development Center, and was also associated with the research staff of the Moore School of Electrical Engineering at the University of Pennsylvania, where he worked on the development of circuits for digital computers, nonlinear techniques for servomechanisms, and other

phases of computer control systems technology. At present he is a member of the faculty of the University of California, Los Angeles.

Dr. Leondes is a member of AIEE, Tau Beta Pi, Sigma Tau, Eta Kappa Nu, and Pi Mu Epsilon.

. . .

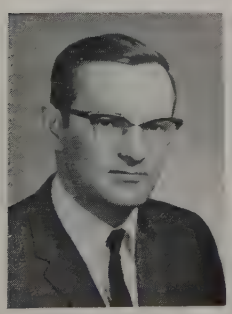
R. H. Loomis was born in Hartford, Conn., on September 4, 1931. He received the B.S. degree in engineering from Trinity College in 1953 and the M.S.E. degree in electrical engineering from The John Hopkins University in 1959.

R. H. LOOMIS

Since 1953, he has been employed by the Air Arm Division of the Westinghouse Electric Corporation in Baltimore, Md., where he has worked on problems associated with the analysis and synthesis of airborne weapons systems. Much of this work has been concerned with the detection and tracking equipment required by interceptors and guided missiles.

. . .

Mihajlo D. Mesarovic was born in Yugoslavia in 1928. He was awarded the M.S. degree in electrical engineering from the University of Belgrade, Yugoslavia, in 1951, and the Doctor of Technical Science degree from the Serbian Academy of Science, Belgrade, in 1955. Since 1953 he has been associated with the Institute "Nikola Tesla," Beograd, where he was formerly head of the Automatic Control Division. In 1956 he was Docent at the University of Belgrade. He attended the School for Advanced Study at M.I.T., Cambridge, in 1958 under a Sloan Foreign Post-Doctoral Fellowship. He is now visiting professor at the Case Institute of Technology, Cleveland, Ohio.

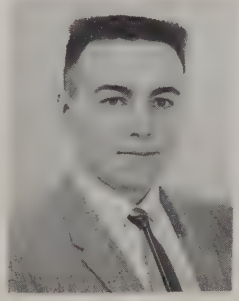


M. D. MESAROVIC

Dr. Mesarovic was a member of the group which initiated the foundation of the International Federation for Automatic Control during the Control Engineering Conference in Heidelberg, Germany, in 1957. He is also an expert member of the Control Committee of the International Federation for High Voltage Power Systems, Paris, and a correspondent member of the Swiss Society for Automatic Control. His articles in the field of automatic control and systems engineering have appeared in American, German, French, and Yugoslavian publications.

. . .

Richard A. Nesbit (S'60) was born on January 17, 1935, in Whittier, Calif. He received the Bachelor of Science in Engineering degree, and the Master of Science in Engineering degree in 1960 from the University of California, Los Angeles. He is presently a graduate student at UCLA and a member of the Adaptive Controls Research Project. He is also on the staff of Space Technology Laboratories, Inc., Los Angeles.

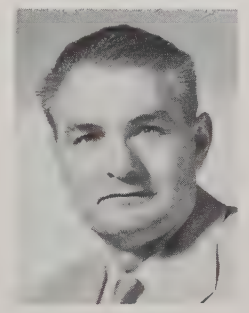


R. A. NESBIT

. . .

George W. Smith (A'54-M'57-SM'60) was born in Westwood, Calif. on December 4, 1923. He received the B.S. degree in physics from the University of San Francisco in 1948, and the M.S. degree in physics from the University of California in 1949.

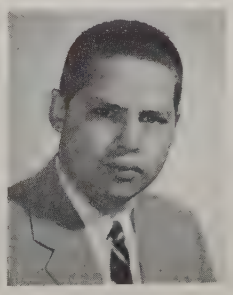
Since 1949 he has been employed in the aircraft industry in the automatic control field. From 1954 to 1959 he was supervisor for systems analysis and computing at the Ryan Aeronautical Company, San Diego, Calif. Since 1959 he has been a research scientist for the Martin Company, Denver Division, with responsibility for adaptive control system studies. He was a lecturer in engineering at UCLA for four years, and is currently a lecturer in engineering at the University of Colorado.



G. W. SMITH

. . .

James E. Vaeth received the B.S. degree in mathematics and physics from Fordham University, New York, N. Y., in 1948 and completed advanced courses in servomechanisms and stability and control. He was a high-school instructor in mathematics and physics for three years before joining The Martin Company, Baltimore, Md., in 1951. He has worked primarily on systems design, stability and control analyses, and error analyses of guidance and control systems for aircraft and missiles. During six years experience in advanced design, he was responsible for autopilot, fire control and inertial guidance systems. He also spent two years in the design and experimental testing of missile launch computers and hydraulic control systems. He was technical director of advanced design projects such as the reconnaissance drone, astronomical satellite, and currently the global surveillance system. He is now design specialist on guidance and control for advanced design projects.



J. E. VAETH

. . .

INSTITUTIONAL LISTINGS

The IRE Professional Group on Automatic Control is grateful for the assistance given by the firms listed below, and invites application for Institutional Listings from other firms interested in the field of Automatic Control.

HUGHES AIRCRAFT CO., Florence and Teale Sts., Culver City, Calif.
Res., Dev., Mfg.: Radar, Space & Commercial Control Systems, Guided Missiles, Solid State Devices

THE RAMO-WOOLDRIDGE CORPORATION
P.O. Box 45215, Airport Station, Los Angeles 45, Calif.

The charge for an Institutional Listing is $75.00 per issue or $125.00 for two consecutive issues. Applications for Institutional Listings and checks (made out to the Institute of Radio Engineers, Inc.) should be sent to Mr. L. G. Cumming, Technical Secretary, Institute of Radio Engineers, Inc., 1 East 79th Street, New York 21, N. Y.